A Companion to Lang's *Algebra*

(Last revised August 25, 2006; based on 3rd Ed., 4th printing of revised Springer version)

notes by George M. Bergman

In these notes, I attempt to bring together in an orderly arrangement materials I have given to my classes over the years when teaching Berkeley's basic graduate algebra course, Math 250, from Lang's *Algebra* – motivations, explanations, supplementary results and examples, advice on material to skip, etc.. I follow this expository material with some exercises not in the text that I particularly like, together with notes on a few of the exercises in the text.

This version of the handout is based on many passes through the first Semester of the course, Math 250A, but only one occasion when I taught 250B from Lang, in Spring 1997. Thus, the coverage of the 250A material is more complete and probably better written than that of the 250B material. Sections of Lang that I have not used in the course are in general not covered.

Order of topics. Math 250A generally treats group theory in some depth, gives some elementary category theory and basic results on rings and modules, and then an extensive development of Galois theory. (Linear algebra, though logically in the same class of fundamental algebraic subjects as the above, is generally postponed till the second Semester, 250B, so that Galois theory can be covered adequately in 250A.) As shown in the list below, in using Lang I follow approximately his ordering of this material, with a few deviations.

Logical Prerequisites (pp.ix-x). Appendix 2.1. §§I.1-6. Appendices 2.2-2.4. §§I.7-12 except §I.10. §§II.1-2. §§III.1-4 except the latter half of §III.2 (from 3d line of p.125). §§II.3-5. §§III.5-6. §§IV.1-6 except §5. §§V.1-6. §§VI.1-9, but ending VI.2 at p.273 and VI.9 at p.299. §VIII.1.

In these notes, for the convenience of those not necessarily following the above order, I have put the commentary on the various sections in the order in which these occur in Lang. However, in cases where the order of the above list deviates from that in Lang, my comments on the sections occurring later in the above list may refer to some sections that occur earlier in that list but later in Lang; e.g., I assume Zorn's Lemma (Lang's Appendix 2.2) in discussing free abelian groups (Lang's §I.7), and I assume basic results on modules, and the concept of an algebra (§§III.1-4) when discussing the ring-theoretic ideas of §§II.3-5.

The material of 250B is less fixed than that of 250A. The additional sections that I covered the one recent time I taught 250B using Lang, and for which there are therefore notes in this Companion, are

§§III.7-9, §X.4, §§XIII.1-7, §§XIV.1-3, §§XVI.1-8.

Pre-comments, post-comments, etc.. Some of my comments are written to be read *before* you read the indicated passage in Lang, to help prepare you for what is done there; others discuss what he has done, assuming you have already read the passage. I will mark these two kinds of comments [<] and [>] respectively. Comments which are to be read right when you reach a certain point in the text will be marked [=]. Finally, the mark [~] means that the order relative to the passage in Lang is not critical.

Long comments. Though most of the items in these notes are short comments on specific points in the text, there are also some longer passages. Some of these develop material which Lang leaves out but which I wished to include in the course (e.g., on structures of groups of small orders), others concern heuristics, or motivate and discuss an open problem. The beginnings and ends of such passages are marked with the symbols ⌐ and ⌙ in the right-hand margin.

Numbering. Results in Lang are referred to here by their chapter number, followed by the section-and-result number they have in that chapter of Lang. E.g., what Lang refers to as Theorem 5.1 of Chapter VI is referred to here as Theorem VI.5.1. Results proved in these notes will be numbered similarly, but with a ''c'' (for ''Companion'') before the final number; e.g., my notes to Lang's §I.6 contain a series of results, numbered Lemma I.6.c1, Lemma I.6.c2, etc.. Labeled displays in this Companion will likewise be numbered (c1), (c2), etc..

The collection of exercises near the end of this Companion is arranged according to the *section* of Lang to which the exercises are related, and numbered accordingly. E.g., the first exercise relating to Lang's §III.2 is numbered III.2:1. In addition, I give notes regarding some of the exercises in Lang. I refer to the

*n*th exercise at the end of Chapter *M* as ''Exercise *M*:L*n*'' (e.g., VI:L8), since Lang does not classify his exercises by section.

Errata. I include, amid the notes to the text and to the exercises, indications of errata that I think are worth bringing to the reader's attention. I also made a practice of sending lists of errata to Lang, and many of these (especially those that are simple typographical errors) were corrected as of the next printing. In these notes I give after the Exercises a list showing the errata to earlier printings of the third edition that have subsequently been corrected (in case you have a copy from one of those printings), as well as errata to the current printing that are not important enough to note below (e.g., misspellings, and symbols in wrong fonts). If any reader discovers errata which are not listed either in the main part of these notes or in that supplementary list, I hope he or she will bring them to my attention.

How do you tell which printing of Lang you have? If you have the Springer version (the ''Revised Third Edition'', with the bright yellow-and-white cover), look at the back of the title page. Four lines from the bottom you will find a line beginning ''9 8 7 ...''. If it goes all the way down to ''1'', you have the *first* printing, if it only goes down to ''2'', the *second*, etc.. If you have the Addison-Wesley version (with a patterned pale blue cover), look at the last line of the back of the title page. If this line begins ''1 2 3 ...'' you have the first printing, if it begins ''2 3 4 ...'' you have the second printing, etc..

On Lang's ''Examples''. Lang frequently gives ''examples'' taken from advanced areas of mathematics which beginning graduate students cannot be expected to be familiar with. Students need not worry if they do not have the background assumed in such examples. I make clear below which they are.

Google Book Searching through Lang. If you want to look for some phrase in Lang and the index doesn't help, it may be useful to go to `http://books.google.com/`, type in

`inauthor:Lang intitle:algebra date:2002-2002`

and then, in the resulting display, click on ''About this book'', and type what you are looking for into the field ''Search in this book''. The results are of limited accuracy – I have run into important errors due to faulty character-recognition on poorly digitized pages; but sometimes the results have been helpful. I'd be interested in hearing whether you find this technique useful.

Additional handouts; web access. There are some handouts that I use in teaching 250A, either expository material or sets of exercises, which are not specific to this text. If I were to incorporate those handouts into the body of these notes, then every time I modified them, I would have to modify both the version in these notes, and the text-independent version, and I probably wouldn't do a good job coordinating the changes. Hence, they constitute separate sheets, which I have included at the end of this Companion. The points where these handouts are to be used will be noted in this Companion. The handouts are as follow. (Note that 2 pp. = one sheet.)

The Axiom of Choice, Zorn's Lemma, and all that, 4 pp.
A principal ideal domain that is not Euclidean, developed as a series of exercises, 1 p.
Lüroth's Theorem and some related results developed as a series of exercises, 2 pp.
Quadratic reciprocity developed from the theory of finite fields as a series of exercises, 2 pp.
Solution in radical of polynomial of degree ≤ 4, developed as a series of exercises, 2 pp.

Some other handouts, which I use in 250B, are not included with this Companion, but can (along with the above) be found through my web page, `http://math.berkeley.edu/~gbergman`.

Other texts. Here are a few books other than Lang that you might like to look at for alternative developments of the same and related material:

Thomas Hungerford, *Algebra* (the other commonly used text for Math 250A)
B. L. van der Waerden, *Modern Algebra* (an old classic)
David S. Dummit and Richard M. Foote, *Abstract Algebra* (more elementary than Lang, with many detailed examples)
P. M. Cohn, *Basic Algebra*, and for some Math 250B material its sequel, *Further Algebra and Applications*.

The next few should also be good, but I haven't had a chance to look closely at them:
Nathan Jacobson, *Basic Algebra*

S. Mac Lane and G. Birkhoff, *Algebra* (not to be confused with G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, an undergraduate text)
I. M. Isaacs, *Algebra, a Graduate Course*
Joseph J. Rotman, *Advanced Modern Algebra.*
N. Bourbaki, *Éléments de Mathématique* (an encyclopedic development), various parts.

For a motivated development of the topics of category theory and universal algebra (a.k.a. ''general algebra''), introduced in §§I.11-12 of Lang, I recommend some course notes I have been developing for Math 245:

G. M. Bergman, *An Invitation to General Algebra and Universal Constructions. 398 pp.*. This can be bought at my office for $40, or from the publisher Henry Helson for $45 by mail, or viewed online via my web page. Below, I will refer to it as ''my Math 245 notes''.

NOTES ON THE TEXT

Re ''Logical prerequisites''. pp.ix-x.

P.*ix*, first two paragraphs [>]: There are three symbols, $\subseteq$, $\subset$, and $\subsetneqq$, one or another pair of which are used by different authors for the two concepts ''is a subset of'' and ''is a proper subset of''. Some authors use $\subseteq$ and $\subset$ respectively, paralleling the use of $\leq$ and $<$, while others (in particular, many Eastern Europeans, and Lang) use $\subset$ for ''is a subset of'' on the grounds that the more basic concept should have the simpler symbol, and then use $\subsetneqq$ (if anything) to explicitly write ''is a proper subset of'' (e.g., in Lang, p.112, second display). To avoid ambiguity, I will (unlike Lang) use $\subseteq$ for ''is a subset of'', while (like Lang) I will use $\subsetneqq$ for ''is a proper subset of'' on the rare occasions when a symbol for proper inclusion is needed.

When f is a map and A a subset of its domain, the notation $f(A)$ for the image of A under f is, strictly speaking, ambiguous, because some subset of the domain might also be an *element* of the domain. Like Lang, we shall use this as a convenient notation; but we will keep in mind that it should be avoided when it could lead to confusion.

Note (if you have taken a course using different notation) that functions will in general be written to the *left* of their arguments, and composed accordingly.

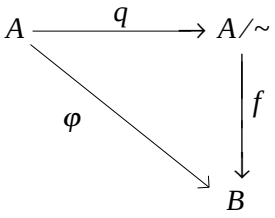
P.*x*, end of page [>]: A *commutative square* is so called because (as a member of the class once elegantly put it), it has the property that ''going to the right commutes with going down''. The term commutativity is extended to diagrams of other shapes by analogy with this case.

In Lang's statement that to verify that a diagram composed of squares and triangles is commutative, ''it suffices to verify'' that each triangle and square in it is commutative, replace ''suffices'' by ''often suffices (depending on the layout of the diagram)''. Cf. Exercise 0:1 in this Companion for more details.

An indexed family of elements, which Lang writes $\{f(i)\}_{i \in I}$, I shall write $(f(i))_{i \in I}$. Lang's notation is misleading, because it looks like set-notation, yet distinct families can have the same sets of elements. (Note, incidentally that the product set $A_1 \times \ldots \times A_n$ referred to by Lang near the bottom of this page is constructed as the set of all families $(a_i)_{i \in \{1, \ldots, n\}}$ with $a_i \in A_i$ for all i.)

The concept of an *equivalence relation* on a set A may be thought of as an abstraction of the properties of the relation $\{(a, a') \mid f(a) = f(a')\}$ induced by a function $f: A \to B$. Its importance grows out of the fact that this relation determines the ''form'' of the image of A under f. Given an equivalence relation $\sim$ on a set A, the quotient set $A/\!\sim$ is a set B with a function $q: A \to B$ which induces (in the above sense) precisely the given equivalence relation $\sim$. Although the ''set of equivalence classes'' construction is a convenient way to prove the existence of such a B, I advise against thinking of the elements of $A/\!\sim$ always as subsets of A; conceptually, I prefer to think of $A/\!\sim$ as ''a set obtained from A by gluing together certain of its elements in a way specified by $\sim$''.

If $f\colon A \to B$ is a map, and $\sim$ the relation $\{(a_1, a_2) \mid f(a_1) = f(a_2)\}$, then there is a natural bijection between $A/\!\sim$ and $f(A) \subseteq B$. But there is a still more useful statement, (ii) below:



Lemma 0.c1. *Suppose $\sim$ is an equivalence relation on a set A. Then the canonical map $q\colon A \to A/\!\sim$ has the properties* (i) $a \sim a' \Leftrightarrow q(a) = q(a')$, *and* (ii) *if $\varphi\colon A \to B$ is any map such that $a \sim a' \Rightarrow \varphi(a) = \varphi(a')$, then there exists a unique map $f\colon A/\!\sim \to B$ such that $\varphi = fq$, i.e., making the diagram at right commute.* □

Statement (ii) is a formalization of the ritual of verifying that the map $(A/\!\sim) \to B$ is ''well defined''. (Lang's discussion of this point does not distinguish clearly between the above ''f'' and ''φ''. It can be corrected by changing the words ''first give its value on an element'' on p.*x*, 9th and 10th lines from bottom, to ''express its value as a function of an element''.)

Lang mentions that Zorn's Lemma will be assumed. This is developed in his Appendix 2.2, and in my handout on ''The Axiom of Choice, Zorn's Lemma, and all that'', which, as indicated in the notes below, is to be read before §I.7.

Chapter I. Groups.

Re §I.1. Monoids.

P.3, bottom [>]: An element e satisfying $ex = x = xe$ for all x is variously called an ''identity element'', a ''unit'', or a ''neutral element''. The first term invites confusion with *identities* satisfied by particular groups or monoids (e.g., the commutative identity), the second is also the term used for any invertible element of a ring, and the last is awkward; so I can only recommend using whatever term has the least disadvantages in a given situation.

A set with an associative operation, but not necessarily having a neutral element, is called a *semigroup*. I actually prefer to call a monoid a ''semigroup with neutral element'', but the term ''monoid'' has become standard, so in these notes I will follow Lang and use it.

Lang will give some examples on pp.6-7, but let me give some quick examples of semigroups and monoids here. (a) *Fundamental example*: If X is a set, then the set of all maps $X \to X$ forms a monoid under composition, with the identity map as neutral element. (b) We can make any set X a semigroup by defining $xy = x$ for all $x, y \in X$. (This is called the ''left zero multiplication''. Setting $xy = y$, we likewise get the ''right zero multiplication''.) (c) If X is a set, we can make the set of all its subsets a monoid, by defining $st = s \cap t$.

How does one formalize the concept of a monoid as a set ''given with'' an operation? If one wants to be precise, one regards the monoid as an ordered pair, $G = (|G|, \mu)$, where $|G|$ is the underlying set of the monoid, and $\mu\colon |G| \times |G| \to |G|$ the operation. This is done in any course on universal algebra (and in my Math 245 notes). Note that where we now write $g \in G$, the above formalization would have us write $g \in |G|$. An alternative preferred by many workers in universal algebra, which allows one to keep writing $g \in G$, is to write $\mathbf{G} = (G, \mu)$. In this course, we simply note that the formalism of ordered pairs (or ordered 3-tuples, since there are good arguments for writing a monoid as $(|G|, \mu, e)$, even though the first two components determine the third) can be used when one needs to make things precise; but we will stick to standard notation here.

P.6, line after first display [>]:

On set-indexed products in commutative monoids.

Let me try to give a more precise and better organized derivation of some consequences of the commutative law. Worked out in detail, it becomes rather lengthy, but I think there is a certain elegance in

the description that develops.

Lemma I.1.c1. *Let G be a commutative monoid. Then there is a unique way of associating to every finite set I, and I-tuple $(x_i)_{i\in I}$ of elements of G, an element of G which we will denote $\prod_I x_i$, in such a way that*

(i) (*Empty product*) *Taking $I = \emptyset$, we have*

$$\prod_{\emptyset} x_i = e.$$

(ii) (*Recursive step*) *If $(x_i)_{i\in I}$ is any I-tuple and j any element of I, we have*

$$\prod_I x_i = (\prod_{I-\{j\}} x_i)\, x_j.$$

Proof. We shall prove inductively that for each n, there is a unique way of defining products of I-tuples of elements of G for those I with $\mathrm{card}(I) \leq n$, such that (i) and (ii) hold.

The case $n = 0$ is clear from (i). So let $n > 0$, and assume by induction that our result is true for all index sets of cardinality $< n$.

Given a set I of cardinality n, the right-hand side of the display in (ii) is defined, by our inductive assumption, for every choice of $j\in I$; we claim that all these choices give the same value. Indeed, suppose j_1 and j_2 are distinct elements of I. Using j_1, we get the value $(\prod_{I-\{j_1\}} x_i)\, x_{j_1}$. Now applying (ii) to the set $I-\{j_1\}$, we see that this value equals $((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_2})\, x_{j_1}$. Similarly, when we use $j = j_2$, we get $((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_1})\, x_{j_2}$. We verify that these are equal using the associative and commutative laws:

$$\begin{aligned}((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_2})\, x_{j_1} &= (\prod_{I-\{j_1,j_2\}} x_i)(x_{j_2}\, x_{j_1}) \\ &= (\prod_{I-\{j_1,j_2\}} x_i)(x_{j_1}\, x_{j_2}) = ((\prod_{I-\{j_1,j_2\}} x_i)\, x_{j_1})\, x_{j_2}.\end{aligned}$$ □

We note two simple special cases, the descriptions of products over 1- and 2-element sets:

$$\prod_{\{i_0\}} x_i = x_{i_0}, \qquad \prod_{\{i_0,\, i_1\}} x_i = x_{i_0}\, x_{i_1}.$$

We also see that whenever I is a set $\{1, \ldots, n\}$, our $\prod_I x_i$ agrees with the meaning given to this symbol for arbitrary monoids on p.4 of Lang. Another result quickly verified by induction is that for any finite set I,

$$\prod_I e = e.$$

The next result describes how, in products defined as above, one can collect terms:

Lemma I.1.c2. *Let $f\colon I \to J$ be a map of finite sets, and $(x_i)_{i\in I}$ an I-tuple of elements of a commutative monoid G. Then*

$$\prod_{i\in I} x_i = \prod_{j\in J} (\prod_{i\in f^{-1}(j)} x_i).$$

Proof. We shall again use induction on $\mathrm{card}(I)$.

When $\mathrm{card}(I) = 0$, point (i) of the preceding lemma shows that the left-hand side of the desired equation is e, and that the right-hand side is $\prod_J e$, which we have noted also equals e.

So now suppose $\mathrm{card}(I) > 0$, and assume inductively that the lemma holds for index sets of smaller cardinality. Take any $i_0 \in I$, and let $j_0 = f(i_0)$. By our recursive definition of $\prod_I x_i$, i.e., condition (ii) of the preceding lemma, we have

$$\prod_{i\in I} x_i = (\prod_{i\in I-\{i_0\}} x_i)\, x_{i_0}.$$

If we now let f' denote the restriction of $f\colon I \to J$ to $I - \{i_0\}$, our inductive hypothesis says we can write the right-hand side of the above equation as $(\prod_{j\in J}(\prod_{i\in f'^{-1}(j)} x_i))\, x_{i_0}$. Again applying condition (ii), this time to the index-set J and its element j_0, this becomes

$$((\prod_{j\in J-\{j_0\}}(\prod_{i\in f'^{-1}(j)} x_i))(\prod_{i\in f'^{-1}(j_0)} x_i))\; x_{i_0}.$$

We can now use associativity to bracket together the terms $(\prod_{i\in f'^{-1}(j_0)} x_i)\, x_{i_0}$, and apply condition (ii) to reduce their product to $\prod_{i\in f^{-1}(j_0)} x_i$. We also note that for $j \neq j_0$, the set $f'^{-1}(j)$ is the same as $f^{-1}(j)$; thus our product equals $(\prod_{j\in J-\{j_0\}}(\prod_{i\in f^{-1}(j)} x_i))(\prod_{i\in f^{-1}(j_0)} x_i)$, which by a final application of (ii) reduces to $\prod_{j\in J}(\prod_{i\in f^{-1}(j)} x_i)$, as claimed. □

Having gotten the above results in the context of finite products, where we could use induction, we may now generalize them to products that are possibly infinite but ''essentially finite'':

Lemma I.1.c3. *There is a unique way of associating to every* (*not necessarily finite*) *set I, and every I-tuple $(x_i)_{i\in I}$ of elements of G almost all of whose terms are e, an element $\prod_I x_i \in G$, in such a way that*

(iii) *For finite I, this value agrees with that defined by Lemma I.1.c1, and*

(iv) *If $(x_i)_{i\in I}$ is such an I-tuple, and I_0 is a subset of I containing all i such that $x_i \neq e$, then $\prod_I x_i = \prod_{I_0} x_i$.*

Moreover, this extended operation still has the properties listed in Lemmas I.1.c1 and I.1.c2.

Sketch of Proof. Given an I-tuple $(x_i)_{i\in I}$ as in the above statement, let us take any *finite* subset $I_0 \subseteq I$ containing all i such that $x_i \neq e$, and define $\prod_I x_i = \prod_{I_0} x_i$. To show that this is well-defined, we want to show that this product is the same whatever I_0 we choose. Let us first consider the case of two choices one of which is contained in the other, $I_0 \subseteq I_1$. Let $f\colon I_1 \to \{0, 1\}$ be defined by $f(i) = 0$ if $i \in I_0$, $f(i) = 1$ otherwise, and let us apply the preceding lemma to this map f. This gives an expression of $\prod_{I_1} x_i$ as the product of two factors, one of which is $\prod_{I_0} x_i$, while the other is a product of e's, which we know equals e. Hence $\prod_{I_1} x_i = \prod_{I_0} x_i$. Now given two sets I_0, I_1 neither of which is assumed contained in the other, we can apply the above case twice to get $\prod_{I_1} x_i = \prod_{(I_1 \cap I_0)} x_i = \prod_{I_0} x_i$, as required.

It is straightforward to verify that the properties we established earlier for finite products carry over to these ''essentially finite'' products, using (iii) and (iv) above. □

We see that the results at the tops of pages 5 and 6 of Lang are particular cases of Lemma I.1.c2 and its generalization to essentially finite products: we get the first by taking $I = J = \{1, \dots, n\}$ and letting f be a permutation; the second by considering a product-set $I \times J$ and its projections onto the two factors I and J.

The concept of homomorphism allows us to state one more important property of set-indexed products, easily proved by induction on $\mathrm{card}(I)$ when I is finite, and using condition (iv) of the preceding lemma in the possibly infinite case:

Lemma I.1.c4. *If $h\colon G \to H$ is a homomorphism of commutative monoids, and $(x_i)_{i\in I}$ a family of elements of G, almost all of which equal e, then*

$$\prod_I h(x_i) \;=\; h(\prod_I x_i). \quad \square$$

Re §I.2. Groups.

P.7, 7th line from bottom [=]: After ''*assume that every element has a left inverse*'', add ''*with respect to* e''.

P.8, after first 3 lines [=]: Lang has proved that one-sided inverses are two-sided inverses under appropriate assumptions. For an example where such conditions are not satisfied, consider the monoid G of maps of $\mathbf{Z}$ into itself, and the element $a \in G$ defined by $a(n) = 2n$; you should be able to find an inverse to a on one side, but show that a has no inverse on the other. Can you find a semigroup containing a one-sided unit that is not a two-sided unit?

P.8, first Example [<]: By $M(S, G)$, Lang means the set of all set-maps from S to G. This is more commonly written G^S.

P.8, second example [>]: Lang has probably included the word ''nonempty'' because some students are confused about maps involving the empty set $\emptyset$, and he doesn't consider it worth the bother to deal with. But if we ignore these problems now, they will take revenge later, so let us get things straight. Recall that a map $f: X \to Y$ is (or at least, is determined by) a set of ordered pairs, $\{(x, f(x)) \mid x \in X\}$. We now note that

if X is a nonempty set, there is no map $X \to \emptyset$, but

for any X, there is a unique map $\emptyset \to X$, namely the map whose graph is the empty set of pairs.

Hence, there is exactly one map $\emptyset \to \emptyset$, and this is a permutation. So we may delete the word ''nonempty'' from the definition of $\mathrm{Perm}(S)$, and note that $\mathrm{Perm}(\emptyset)$ is a one-element group.

P.8, ''**Example. The group of automorphisms**'' [<]: I *don't* advise following Lang's recommendation unless you are already familiar with category theory. If you are not, skip this ''example''.

P.9, line 2 [>]: The term ''cyclic'' is natural for this class of groups because of the form the finite ones have. By extension, the term is also frequently used to mean ''generated by a single element'' when referring to any sort of mathematical object, even if the forms of these objects are not ''cycle''-like.

It is interesting to examine the structures of *cyclic monoids*. Their description is not quite as simple as that of cyclic groups, but not too hard to work out. The details can be entertaining to develop in class, or you might like to work them out for yourself.

P.10, top paragraph [>]: Logically, the definition of a homomorphism of monoids belongs in the preceding section. Note that such a homomorphism is required not only to respect multiplication, but to take unit element to unit element. A map $G \to G'$ respecting multiplication will clearly send e_G to an element i of G', such that $i^2 = i$. Such an element is called an *idempotent*. In a *group*, it is easy to see that the only idempotent is the neutral element, so in the definition of a group homomorphism, many authors leave out the condition of respecting neutral elements. But a monoid can have many idempotent elements (e.g., in the monoid of subsets of a set, under the operation of intersection, all elements are idempotent), so the condition of respecting neutral elements is an essential part of the definition of a monoid homomorphism.

P.10, middle [>]: ''The existence of an isomorphism between two groups G and G' is sometimes denoted $G \approx G'$.'' Some other common notations for this relation are $G \cong G'$ and $G \simeq G'$.

P.10, italicized statement at the bottom of page [>]: Here are two ways of proving this. In each case, let us assume (modifying Lang's notation) that $p, q: G \to H$ are homomorphisms, both of which, when restricted to the generating set S, give the map $f: S \to H$. We wish to prove $p = q$.

Sketch of naive proof. Since G is generated by S, every element $g \in G$ may be written as a product of elements of S and their inverses. Use the fact that p and q respect composition and inverses to write $p(x)$ and $q(x)$ as expressions in the images of elements of S under f. We get the same expressions, hence $p(x)$ and $q(x)$ have the same value.

Good proof. It is immediate to verify that the set $A = \{x \in G \mid p(x) = q(x)\}$ is a subgroup of G, and contains S. Hence it contains the *least* subgroup containing S, i.e., the subgroup generated by S, which in this situation is assumed to be G. So A contains G, i.e., $p = q$.

The characterization used above of the subgroup of a group G generated by a set S, as the *least* subgroup containing S, is an example of what is called a *universal* characterization of a mathematical object, since it describes the subgroup in question in terms of its relation to *all* subgroups containing S. Universal characterizations sometimes yield much more direct proofs than do ''descriptions'' of the objects so characterized.

P.11, parenthetical remark containing first display [=]: This just repeats the proof of the first display on the preceding page.

P.11, italicized statement in middle of page [=]: For ''*a homomorphism*'' read ''*a group homomorphism*''.

P.11, Proposition I.2.1 [>]: This result may be looked at as the *converse* to an important observation: that if one forms the direct product G of two groups H and K, then H and K can be identified with subgroups of G (namely, $\{(h,e) \mid h \in H\}$ and $\{(e,k) \mid k \in K\}$), and that these two subgroups have trivial intersection, commute elementwise, and generate the whole group.

It is often convenient to state the conditions described in the conclusion of the above proposition, as saying that ''G decomposes as the direct product of the subgroups H and K''. But one should bear in mind that there is a subtle difference between the senses of ''being a direct product'' in this situation and as originally defined, since this G does not literally consist of ordered pairs of elements of H and of K (though the proof is based on associating to each $g \in G$ such a pair, namely the unique (h,k) such that $g = hk$); and inversely, in a direct product in the original sense, the given groups are not subgroups of G (though, as noted, they can be naturally identified with such subgroups). When one needs to distinguish these usages, one may say that a group G obtained by the original direct product construction is the ''external direct product'' of H and K, while in the situation of the proposition, G is the ''internal direct product'' of its subgroups H and K. When we come to category theory, we shall discover a unified concept of direct product which includes both of these.

P.12, middle, definition of ''coset representative'' [>]: An important concept (used in the next proposition) is that of ''a set of coset representatives''. This means a subset of G containing *exactly one* representative of each coset of H.

P.12, Proposition I.2.2 [=]: Lang asks us to think of the index-function $(G:H)$ as taking as values the finite cardinals and a nonspecific ''∞''. But as one can easily see from the second paragraph of this proposition, if one defines $(G:H)$ as the cardinality of the set of cosets of H in G, the displayed equation holds as a precise arithmetic statement about these cardinals, whether they are finite or infinite. (So, for instance, one can deduce that an uncountable group can't have a countable subgroup of countable index.)

P.13, end of §I.2 [>]:

Alias and alibi.

To avoid confusion in the study of groups of permutations, it is useful to be aware of what Hermann Weyl called ''the problem of alias and alibi''. As an example, suppose we are studying the permutation group on 5 elements, and we think of these permutations as corresponding to ways of putting marbles labeled $1, \ldots, 5$ into holes labeled $1, \ldots, 5$. Suppose we have a particular such permutation, say (23) (represented by the arrangement with marble 3 in hole 2, marble 2 in hole 3, and all other marbles in the holes with the same numbers), and we wish to describe the effect of following it by the permutation (12). Do we get this by interchanging the marbles that are currently in holes 1 and 2, or by interchanging the marbles with numbers 1 and 2? In the first case, we are letting S_5 act by *alibi* (Latin for ''elsewhere'', i.e., we are letting S_5 act by permuting the 5 *positions*); in the second, it acts by *alias* (Latin for ''otherwise''). For another example: consider the symmetry group of a square drawn on the blackboard. We might imagine a celluloid copy of this square made (say labeled *ABCD*, with *A* at the upper left), and we might then represent elements of this group as ways of placing this copy over the original square. But then, should the group contain operations such as ''flipping about the vertical axis of

the blackboard square'' or ''flipping so as to interchange A with B, and C with D, in the celluloid copy''? If we start with the celluloid square in its original position, these have the same effect, but if we compose symmetries, so that one operation acts after another has moved the celluloid square, we in general get different results.

In these examples, note that what we have been calling ''permutations'' are really bijections between two sets, X (a set of ''places''), and Y (a set of objects), which are given us with some ''standard'' bijection between them. Our symmetry group can be regarded either as a group of permutations of X, or as a group of permutations of Y. Given a bijection $f: X \to Y$, the action by ''alias'' is that of composing f on the left with permutations of Y, and the action by ''alibi'' is that of composing f on the right with permutations of X. If we *identify* X with Y, using our ''standard'' bijection – so that the set of allowed bijections, and the symmetry group of X, and the symmetry group of Y all become the same set – then the action of an element g on the symmetry group by *alias* becomes the ''left translation'' map $x \mapsto gx$ on that group, and the action by *alibi* becomes the ''right translation'' map, $x \mapsto xg$.

Re §I.3. Normal subgroups.

P.14, top [~]: What general computational principle allows Lang to argue (in the first paragraph) that the condition $xH \subset Hx$ is equivalent to $xHx^{-1} \subset H$ and (in the next paragraph) to write $xHyH = xyHH = xyH$ if H is a normal subgroup? The key observation is that the *subsets* of a group or monoid G form a monoid under the multiplication of subsets defined on p.6:

$$SS' = \{xy \mid x \in S,\ y \in S'\}.$$

(Associativity of this operation is an immediate consequence of the associativity of G. The ''unit'' is the singleton $\{e\}$.) In this monoid, singletons clearly multiply like the corresponding elements: $\{x_0\}\{x_1\} = \{x_0x_1\}$; also recall that we abbreviate the product $\{x\}S$ or $S\{x\}$ of a singleton $\{x\}$ and a set S as xS, respectively Sx.

Thus, given $xH \subset Hx$, we can multiply on the right by x^{-1} (i.e., $\{x^{-1}\}$) and get $xHx^{-1} \subset He = H$. You should be able to write down a similar argument showing that if H is a normal subgroup, then it ''commutes'' with all elements. Thus, we get $xHyH = xyHH = xyH$.

P.14, middle [<]: *Motivation for the construction of the factor group.* We noted in our comments on pp.*ix-x* that given a *set* G, to form a picture of the image of a map f with domain G it suffices to know the equivalence relation on G that describes which elements fall together under f. If G is a *group* and f a group homomorphism, we can then describe the *group structure* of the image, using the law $f(x)f(y) = f(xy)$; so the group structure of G and the above equivalence relation completely determine the structure of the group $f(G)$.

Also note that $f(x) = f(y)$ if and only if $f(xy^{-1}) = e$, hence to determine the equivalence relation, it suffices to know the set of elements which have image e. This set turns out to be a subgroup N, and to satisfy $y^{-1}Ny = N$ for all $y \in G$; a subgroup with these properties is called *normal*, and the normal subgroup of G associated in this way with a homomorphism f is called the *kernel* of f. Conversely, if N is a normal subgroup of G, then the quotient-set of G by the equivalence relation $\{(x,y) \mid xy^{-1} \in N\}$ can be made a group G/N in such a way that the quotient map $G \to G/N$ is a homomorphism having N as its kernel. These considerations show that homomorphic images of G are in one-to-one correspondence with normal subgroups of G. (To be more precise, suppose that given homomorphisms of a group G into two groups H_1 and H_2, we define an ''isomorphism of groups with homomorphisms of G into them'' to mean a group isomorphism $H_1 \to H_2$ making a commuting triangle with the given homomorphisms from G. Then normal subgroups of G correspond bijectively to homomorphic images of G, up to isomorphism as groups with homomorphisms of G to them.)

For monoids, on the other hand, the set of elements that a homomorphism sends to the neutral element does *not* determine the structure of the image; hence one must look at the whole equivalence relation to

find that structure. Such an equivalence relation is called a *congruence* on a monoid, and one can write down necessary and sufficient conditions on an equivalence relation on a monoid for it to be a congruence.

Group-theorists generally write $H < G$ to mean ''H is a subgroup of G'', and $H \lhd G$ for ''H is a normal subgroup of G'' (and likewise $G > H$ or $G \rhd H$ if they want to name the larger group first). We shall often use this notation.

Given any subset S of a group G, there is a *least* normal subgroup of G containing S. A few basic facts about this construction are noted in Exercise I.3:4. I recommend looking at that exercise and making sure you see why the assertions are true.

P.14, middle, the ''canonical map'' [>]: A ''canonical'' object means one determined in a special way by the data being considered.

Something you might find it interesting to think about: Suppose G is a group, H a subgroup, and S a set of left coset representatives of H in G. If H is normal, we know that left cosets are the same as right cosets, so these will also be a system of right coset representatives. If H is not normal, we may ask whether we can at least find *some* set of left coset representatives which is also a set of right coset representatives. You might consider both case where H has finite index in G and the general case.

P.15, three italicized sentences [=]: The first and last of these statements are fairly obvious (and, in fact, say the same thing); the middle one takes more thought, but is important. You should check that you can prove it.

P.15, final display [=]: Here ''0'' denotes the trivial group. That is a common symbol for it; ''1'' is often used when G is written multiplicatively. The explicit symbols $\{e\}$, $\{0\}$ and $\{1\}$ can also be used. (The notation ''1'' is implicit in the notation $(G : 1)$ for the order of G.)

P.16, the commutative diagram [<]: You do not have to learn all the diagrammatic results proved over the next three pages; what you should concentrate on is understanding *how* they are proved. You should, however, learn the two boxed isomorphisms on p.17, which in most texts are stated as ''Isomorphism Theorems''. The ideas behind these are discussed below.

P.16, point (**ii**) [~]: Here we might just as well replace the subgroup H by an arbitrary subset S. The result obtained can be thought of as saying that the group G/N is *universal* among groups E given with homomorphisms $G \to E$ having S in the kernel.

P.17, (**iii**) and (**iv**) [>]:

Canonical isomorphisms.

The isomorphisms in the two boxes on p.17 relate subgroups and homomorphic images of a group G and of a factor-group thereof. An important fact, easily verified, but not mentioned by Lang, which will help us understand these results, is:

(c1) *Let $q\colon G \to G/N$ be the canonical map from a group to a factor-group. Then there is a bijective correspondence between subgroups of G/N and subgroups of G that contain N, the map from the former to the latter being given by $H \mapsto q^{-1}(H)$; the map the other way by $H \mapsto H/N = q(H)$. Moreover, H/N is normal in G/N if and only if H is normal in G.*

The first boxed isomorphism on p.17 can now be understood as follows: Suppose we have a group G, a homomorphic image of that group, and a homomorphic image of this homomorphic image. Thus, we can write the first of these images as G/K, where $K \lhd G$; and by the above observation, the second will have the form $(G/K)/(H/K)$ for some normal subgroup $H \lhd G$ containing K. Now we may also regard this group directly as a homomorphic image of G, i.e., its factor-group by the kernel of the composite map $G \to G/K \to (G/K)/(H/K)$. This kernel is clearly H, giving the asserted isomorphism $(G/K)/(H/K) \cong G/H$.

This result can be looked at in still another way: Given a group G and two normal subgroups K and H, with the former contained in the latter, the canonical map $G \to G/H$ has K within its kernel, hence

induces a map $G/K \to G/H$, as in the triangular diagram in point (ii) on p.16. This expresses G/H as a homomorphic image of G/K, equivalently, as a factor-group thereof. The factor-group by what subgroup? H/K!

Note that the final statement of (c1) above concerns this same situation; but the boxed isomorphism tells us more than that statement, namely that corresponding normal subgroups of G and G/N have isomorphic factor-groups.

To motivate the other boxed isomorphism on p.17, again consider the canonical map from G to a factor-group G/K, and now let H be an arbitrary subgroup of G, and let us look for a description of the image of H in G/K under that map. On the one hand, this will be isomorphic to the factor-group of H by the subgroup of its elements that go to the identity element under the map; i.e., to $H/(H \cap K)$. On the other hand, it is a subgroup of G/K, hence has the form L/K, where L is the inverse image of this subgroup in G. This inverse image consists of those elements of G whose images in G/K are equal to images of elements of H. We easily verify that this says $L = HK$; so identifying our two descriptions of the image of H in G/K, we have $H/(H \cap K) \cong HK/K$.

(Lang actually assumes slightly less than we did in the above paragraph: where we assumed K normal, he just says that H lies in the normalizer of K. However, if we let G_0 denote that normalizer, then his assertion follows from our observations applied with the group G_0 in place of G.)

Important remark: Most often, a statement that something is ''isomorphic to'' something else is an incomplete statement. The full statement, which is frequently what one really uses when one applies the result, is that *a particular map* from one object to the other is an isomorphism; or that there is a *unique* isomorphism from one object to the other that *has a certain property*. This is indeed true in the above two cases. The first result should be formulated as saying that there is a unique isomorphism between $(G/K)/(H/K)$ and G/H that makes a commuting triangle with the canonical maps of G into these two groups – briefly, a unique ''isomorphism as images of G''; the second says, similarly, that there is a unique isomorphism between $H/(H \cap K)$ and HK/K as images of H.

P.18, definition of normal tower [>]: Though each G_i is assumed normal in the subgroup immediately above it in the tower, that does not mean it will be normal in *all* higher members of the tower: A normal subgroup of a normal subgroup need not be normal.

(However, if one defines a *characteristic subgroup* to be a subgroup carried into itself by *all* automorphisms of the group, then it is easy to see that a characteristic subgroup of a characteristic subgroup is characteristic. I leave it to you to think of what can be said of a characteristic subgroup of a normal subgroup, or a normal subgroup of a characteristic subgroup.)

P.18, proof of Proposition I.3.1 [<]: Before reading this, note the following

Observations on normal towers.

A key fact used in connection with normal towers is: If $G \rhd N$ are a group and a normal subgroup, and we have a normal tower

$$G = H_1 \rhd H_2 \rhd \ldots \rhd H_m = N,$$

then

$$(H_1/N) \rhd (H_2/N) \rhd \ldots \rhd (H_m/N) = \{e\}$$

is a normal tower for G/N, and the successive factor-groups of this tower are isomorphic to those of the first tower via the first boxed isomorphism on p.17. Moreover, this construction gives a *bijective* correspondence between normal towers starting with G and ending with N, and normal towers starting with G/N and ending with $\{e\}$.

In particular, suppose we look at any pair of successive terms $G_i \rhd G_{i+1}$ in a normal tower for some group G, and ask what sequences of terms may come between G_i and G_{i+1} in refinements of that

tower. The above result shows that the possibilities correspond naturally to the normal towers that start with G_i/G_{i+1} and end with $\{e\}$.

This is used implicitly in the first sentence of the proof of Proposition I.3.1 to pass from a tower for the abelian group G_i/G_{i+1} to a tower between G_i and G_{i+1}, and again in the proof of Theorem I.3.5, where it is combined with the observation that a normal tower for a simple group merely has one factor isomorphic to the group and all other factors trivial.

P.18, fifth line of the proof of Proposition I.3.1 [=]: After ''tower in G''' add ''ending in $\{e\}$''.

P.19, proof of Theorem I.3.2 [<]: The proof that if G is solvable then G/H is solvable is essentially given on the preceding page, in the paragraph beginning ''Let $f: G \to G'$ be a homomorphism''.

P.20, above first example, definition of *simple group* [=]: This definition is a sensible one, but the usual usage of finite group-theorists is different: they call a group simple if it satisfies the property stated by Lang *and* is non-abelian. The reason is that the abelian simple groups, i.e., the cyclic groups of prime order, though straightforward in their structure, are quite different in properties from the other simple groups. In particular, note that for every group G, the subgroup G^c is normal, hence if G is simple (in Lang's sense), G^c is either G or $\{e\}$. The abelian simple groups are those for which G^c is $\{e\}$, while for all others it is G.

P.20, last two sentences of second paragraph of second example [>]: A group theorist tells me that the one known gap in the proof, the ''quasi-thin'' case, was filled in two papers totaling about 1,300 pages by Michael Aschbacher and Steve Smith, published in 2004.

Since the proof of the classification constitutes many thousands of pages of articles, there is still a possibility of hidden errors. People in the field have been working on simplifying parts of the proof, so the chance of undiscovered errors should decrease with time.

P.20, Lemma I.3.3 [=]:

Idea of the Butterfly Lemma

If G is a group, and U/u, V/v are homomorphic images of subgroups of G, one would like to describe the extent to which one can ''relate'' part of the structure of U/u and part of the structure of V/v, based on their common origin in G. If one takes the subgroup of U/u consisting of all elements which are also images of elements of V, one can write this $u(U\cap V)/u$. The analogous subgroup of V/v is $(U\cap V)v/v$. To get a common homomorphic image of these, we must divide each by the subgroup of those elements that are annihilated in the construction of the other. These subgroups are the denominators in the expressions at the bottom of p.20. The Butterfly Lemma says that after making these adjustments, we do get isomorphic groups, the ''common heritage'' of U/u and V/v.

Group-theorists call a *factor-group* of a *subgroup* of a group G a *subfactor* of G. Thus, given two subfactors U/u and V/v of a group G, the Butterfly Lemma characterizes their largest natural ''common subfactor''.

There are a couple of points in the proof of that lemma that Lang skips over. First, why are the four sets $u(U\cap V)$ etc. subgroups? Note that u is normal in U, so $U\cap V$ is contained in the normalizer of u. From this we see that $u(U\cap V)$ is a subgroup; the other cases are seen in the same way. Second, where Lang says of the diagram on p.21 that ''The intersection of two line segments going downward represents the intersection of the groups. Two lines going upwards meet in a point which represents the product of two subgroups'', this gives four descriptions of the one unlabeled point of the diagram; so these four groups must be shown equal! This is not hard once the task is recognized; cf. Exercise I.3:3.

Re §I.4. Cyclic groups.

P.23, start of §I.4 [<]: In §I.2, a ''generator'' of a group or monoid G meant a member of whatever generating set was under discussion – a relative rather than an absolute concept. But in this section, an

element x is called a ''generator'' of G if the singleton $\{x\}$ is a generating set, i.e., if x is what is called a ''cyclic generator'' on p.8. Both usages are common in group-theoretic writing, so one needs to watch out which sense is meant when the term is used.

What Lang calls the *period* of an element of a group (p.23) is usually called its *order*. This concept is most useful when one wants to know ''all about'' some element of a particular group; the concept of an *exponent* of an element is more useful in general arguments, where one has limited information to work with.

P.25, end of proof of (**v**) and whole proof of (**vi**) [=]: Lang relies on results not yet proved. Here are self-contained arguments (though in the proof of (**vi**), I will leave two steps for you to justify):

In (**v**), having proved that the kernel of the map $\mathbf{Z} \to A \times B$ is $mn\mathbf{Z}$, we conclude that the image has order mn, which is the order of the whole group $A \times B$, so the map is surjective.

In proving (**vi**), let us use the group-theorists' notation ''$\langle \ldots \rangle$'' for ''subgroup generated by''; and let us also write our group operation additively, so that we can write things like nG for $\{nx \mid x \in G\}$. Let H be a subgroup of G *minimal* for the property of being noncyclic (i.e., such that $H \leq G$ is noncyclic, but all proper subgroups of H are cyclic. Such an H must exist because G is noncyclic and has only finitely many subgroups.) Now since multiplication by the integer $(H:1)$ is the trivial endomorphism of H, and this map is a composite of multiplications by primes, there must be some prime p such that multiplication by p is not an automorphism of H. Hence pH is properly smaller than H, hence by assumption it is cyclic. Let us write a generator of this subgroup as px, where $x \in H$. Since H is not cyclic, it properly contains $\langle x \rangle$, so we can choose a $y \in H$ not in $\langle x \rangle$. Since $py \in pH = \langle px \rangle$, we can write $py = mpx$ for some integer m. Let $z = y - mx$. Then $z \notin \langle x \rangle$ and $pz = 0$. There are now two cases:

Case 1: x has order divisible by p. Then $\langle x \rangle$ has an element w of order p (by (iv)), and you should be able to show that the subgroup of H generated by w and z is isomorphic to the direct product of two cyclic groups of order p, as desired.

Case 2: x has order not divisible by p. Then it is easy to verify that the element $x' = x + z$ has order divisible by p, and it clearly also has the property we assumed earlier for x, namely that $px' = px$ generates pH. Hence we can repeat the above argument with x' in place of x, and this time we end up in Case 1, and so get the desired conclusion.

Re §I.5. Operations of a group on a set.

P.26, end of top paragraph [=]:

General observation on the source of group and monoid theory.

The concept of group (respectively monoid) arose from consideration of the properties of the automorphisms (respectively endomorphisms) of a mathematical object. The concept of an *abstract group* was a great advance, allowing mathematicians to consider groups without reference to specific permutation representations. But group theory does not cut itself off from its roots – the concept of an action of a group by permutations, introduced in this section, allows us to work with abstract groups, and yet represent them, when convenient, by permutations of sets; and often provides good intuition in studying abstract group theory.

Of course, to know whether permutation representations of groups are really general enough to embrace all of abstract group theory, we may ask whether *every* abstract group G admits a representation as a group of permutations of some set X. Many of you have seen the answer in your undergraduate courses, but let us analyze the question pretending that we haven't. Such a representation of G would correspond to giving a G-set X, such that distinct elements of G have distinct actions on X. (A G-set X with this property is said to be *faithful*.) We observe that if X is any G-set, and x any element of X, then X must contain elements gx for all $g \in G$, and we can describe how G acts on the set of such elements – namely,

$$g'(gx) = (g'g)x.$$

Hence, let us try to construct a G-set by choosing a symbol x, letting X consist of all symbols gx as g ranges over G, and defining the action of G on X as above. It is easy to verify that this does make X a G-set, and that distinct elements of G act by distinct permutations. This gives the desired result, *Cayley's Theorem*, that every group is isomorphic to a group of permutations of a set. We may make a symbolic simplification: since the same symbol x appears at the end of all the elements of X, it provides no information, so we may instead take for our X the set of elements of G itself. This gives the standard proof of Cayley's Theorem, by the observation that the *translation representation* of G (which Lang will introduce at the bottom of this page) is faithful.

P.26, section on conjugation [>]: We can get perspective on the concept of *conjugation* in groups by beginning outside the group-theoretic context, and looking at a bijection

$$\alpha : X \to Y$$

from one set to another. This can clearly be used to ''carry'' any sort of structure on X over to Y. For instance, if we have an I-tuple of elements of X, $(x_i)_{i \in I}$, we get an I-tuple of elements of Y, $(\alpha(x_i))_{i \in I}$. If we have a real-valued function $h: X \to \mathbf{R}$, we find that the corresponding real-valued function on Y is $h\alpha^{-1}$. (How do we see this? Given any $y \in Y$, to find the value at y of the function on Y ''corresponding to h'', we look for the element of x that corresponds to y, namely $\alpha^{-1}(y)$, evaluate h there, and conclude that the result, $h(\alpha^{-1}(y))$, is the value we want.) Now given a map $f: X \to X$, we can likewise get a ''corresponding'' map $Y \to Y$; you should verify for yourself that this is $\alpha f \alpha^{-1}$. Note that this new map will be bijective if and only if f is. If we now turn to the case where X and Y are the same set X, and α and f are two permutations of X, we see that $\alpha f \alpha^{-1}$ still represents the operation on X which ''corresponds to f, under the transformation α of X into itself''.

Returning to the case of two sets X and Y for clarity, we can see intuitively that this operation of ''carrying a permutation of X to the corresponding permutation of Y'' must respect *composition of permutations*; hence in particular, this is true in the same-set case. Since every group can in fact be represented by permutations of a set, every group will have the property that conjugation by a fixed element respects composition, hence is an automorphism of the permutation group.

Finally, if in our ''different-sets'' situation, we have in addition to the bijection α between X and Y, a bijection β between Y and a set Z, we can see that carrying an endomap f of X to the endomap of Y to which it corresponds under α, and carrying this in turn to the endomap of Z to which it corresponds under β, must give the same result as carrying f directly to the endomap of Z to which it corresponds under $\beta\alpha$. Thus, a composite of conjugation maps gives conjugation by the composite map. In particular, for a group G, the construction sending each element α to the operation of conjugation by α is a *group homomorphism* $G \to \mathrm{Aut}(G)$.

All the results obtained above can be verified *computationally* in less space than the above discussion! The point of this exposition is to give us intuition on the subject. In particular, it often allows us to ''see'' without computation the *form* that the conjugate of one element by another in a group of permutations of a structure will take. For instance, in the symmetric group S_5 (which Lang will define on p.30, but most of you saw in your undergraduate courses), if we take the permutation acting by $1 \mapsto 2 \mapsto 3 \mapsto 1$ – in the notation Lang introduces on p.30, [123] – and conjugate it by the permutation [25], we must get [153].

If you haven't seen the symmetric group, you should read the short paragraph defining it on p.30 now, then return to p.26, since Lang will assume familiarity with it at the bottom of p.28.

P.26, bottom [=]: The maps T_x that Lang defines are more precisely called *left translations*. One also speaks of the *right translation* maps R_x given by $R_x(y) = yx$.

The map $x \mapsto R_x$ is not a homomorphism $G \to \mathrm{Perm}(G)$; rather, it satisfies $R_x R_y = R_{yx}$. Thus, these maps do not define an action of G on itself; nevertheless it is of some importance, so it is best to always speak of ''left translation'' and ''right translation'', rather than just ''translation''.

(In fact, $x \mapsto R_x^{-1}$ does define an action of G on itself. Moreover, every map T_x commutes with every map R_y: $T_x R_y = R_y T_x$. One can deduce from this and the fact that $x \mapsto R_x^{-1}$ gives an action of G on itself that the map $x \mapsto T_x R_x^{-1}$ also gives such an action. Namely?)

P.27, paragraph just below definition of **isotropy** group [<]: If you are reading this in my 250A or another course that has you read this section before §I.3, then simply understand that the ''normalizer'' mentioned in this paragraph is something you will learn about very soon. (Or better – work out for yourself what the isotropy subgroup of an element under the conjugation action is, then look up ''normalizer'' in the index and see whether it agrees with what you have gotten.)

P.28, first sentence of last paragraph [<]: S_n denotes the group which, in the notation of the second example of p.8, would be written $\mathrm{Perm}(\{1, \dots, n\})$. (Lang will introduce the notation S_n on p.30.)

The remaining sentences of this paragraph refer to subjects you should not expect to be familiar with.

P.28, bottom [>]:

Notes on group actions.

I have pointed out that the essential motivation for the concept of *group* is that it embodies the properties of the collection of all permutations of a set, or more generally, of the collection of all automorphisms of a mathematical object. Thus, to get intuition on some statement about groups, it is useful to think of what the statement means in the case of a group of permutations. What about the concept of a subgroup of a group G?

The most useful general source for that concept, I claim, is the construction of the *isotropy subgroup* G_x of an element x of a G-set X (also often called the *stabilizer* of x). Indeed, most of our examples of subgroups really arise in this way. Consider, for example, the symmetry group of the square, D_4, which is the subgroup of the permutation group S_4 of the vertices of a square comprising those permutations that preserve all distances between vertices. If we let X denote the set of all *metrics* on the 4-element set consisting of the vertices, $\{1, 2, 3, 4\}$, then S_4 acts on X: a permutation σ takes each metric $d(\ ,\ )$ to the metric $d(\sigma^{-1}(\ ), \sigma^{-1}(\ ))$. The definition of D_4 can now be seen to say that, if we let d denote the metric on $\{1, 2, 3, 4\}$ such that $d(i, j)$ is the distance between the ith and jth vertices of the unit square, then D_4 is the *isotropy subgroup* $(S_4)_d$ under this action of S_4 on metrics. Other descriptions of subgroups of various groups, as comprising those permutations that ''preserve'' certain structure on a certain object, translate similarly to characterizations of them as isotropy subgroups. (In considering particular cases, there is no need to always translate such definitions of subgroups into the language of isotropy subgroups. But having seen that they can be so translated gives us confidence in using the isotropy-subgroup concept as a general way to look at subgroups.)

Given a group G and a subgroup H, does there always exist a G-set X and an element $x \in X$ with $G_x = H$? Yes: as shown in exercise I.5:1, G/H is a G-set with such an element. That result could have been motivated in the same way as the proof of Cayley's Theorem above.

Some examples of the translation between group-theoretic concepts and G-set concepts: If A and B are isotropy subgroups of elements of two G-sets, $x \in X$ and $y \in Y$ respectively, then their *intersection*, $A \cap B$, is the isotropy subgroup of the element (x, y) of the G-set $X \times Y$. If H is a subgroup of a group G, say the isotropy subgroup of an element x of a G-set X, then the *conjugate* gHg^{-1} of H by an element g can be interpreted as the isotropy subgroup of the translate gx of x. (This fits in with the viewpoint that conjugation of g by α means taking ''the permutation corresponding to g under a shift of location given by α''.)

It follows that a subgroup H of a group of permutations G of a set X will be *normal* in G if H can be described in terms of some *G-invariant* structure on X.

We can use this observation to get some insight to the fact noted in our discussion of normal towers, that a normal subgroup of a normal subgroup need not be normal. E.g., in the group G of symmetries of

the square, the subgroup N consisting of the four elements preserving both the horizontal and the vertical axes is normal, and within N the subgroup M generated by reflection through the vertical axis is normal, but M is not normal in G. These facts, easily verified computationally, can be interpreted conceptually by noting that N can be characterized in terms of structure on the square (namely, the set consisting of the horizontal and vertical axes) which is G-invariant, and M is normal in N because it can be characterized in terms of a structure (namely, the vertical axis) which is N-invariant; but, since the latter structure is not G-invariant, M may not be (and indeed, is not) normal in G.

If G is any group, then a group which, like the M of the above example, appears in a normal tower for G is called a *subnormal subgroup* of G. If, in this example, we replace the square by a regular 2^n-gon, one can show that the subgroup generated by reflection about the vertical axis occurs in a normal tower of height n, but not in any normal tower of smaller height.

P.30, last display [=]: Can you see that the first two equalities are correct? The difficulty one typically has in applying the definition Lang has given is a classic example of ''alias vs. alibi'' confusion (cf. comment in this Companion re p.13 of Lang).

Incidentally, on the next page Lang will take for f a polynomial function. If we regard polynomial functions as induced by members of the ring of polynomials over $\mathbf{Z}$ in n indeterminates, $X_1, \dots, X_n$ (to be introduced formally in the next chapter), then the corresponding action of S_n on this polynomial ring is by permutation of the indeterminates, $\pi(\sigma)(X_i) = X_{\sigma(i)}$; and in this case we have $\pi(\sigma)\pi(\tau)\, f(X_1, \dots, X_n) = \pi(\sigma)\, f(X_{\tau(1)}, \dots, X_{\tau(n)})$, which nevertheless leads to $\pi(\sigma)\pi(\tau) = \pi(\sigma\tau)$. I leave it to you to ponder this paradox.

P.32, line after first display [=]: For H_v/H_{v-1} read H_{v-1}/H_v.

P.32, Proof of Theorem I.5.5 [<]: I recommend not trying to follow the details, but working out for yourself that if N is any normal subgroup of S_n $(n \geq 5)$ and g a nonidentity element of N, then by playing with conjugates of g, you can come up with an element that moves few points, from that get a 3-cycle, and from this get the general element of A_n.

(Incidentally, in the first paragraph of Lang's proof, every printing through the current one has some confusion regarding the conditions on σ. In the current printing, there is merely the unnecessary phrase ''or the identity'' four lines from the bottom, though σ has been assumed a nonidentity element; the 3rd printing likewise had that phrase plus a following unnecessary sentence. The 4th and 5th printings, more seriously, left out the assumption on an earlier line that $\sigma \neq 1$, though without this, the maximality condition makes $\sigma = 1$, and one can deduce nothing about 3-cycles.)

P.33, end of §I.5 [>]: The material below, not in Lang, is required for my 250A class. It will be assumed in the analysis of groups of order pq a few pages from now.

Semidirect products.

When a turtle *retracts* into its shell, the shell stays fixed, and the turtle's legs, head and tail move into it. Similarly, a *retraction* of a group G onto a subgroup H means an endomorphism $r\colon G \to G$ such that r fixes all elements of H, and carries all elements of G into H. If $r\colon G \to G$ is such a retraction, and N its kernel, it is easy to verify that every element $g \in G$ can be written uniquely as a product of a member of H (namely, $r(g)$) and a member of N (namely, $r(g)^{-1}g$). To describe how such products are multiplied, we need to know the action of H on N by conjugation: If for fixed $h \in H$ we write $\psi(h)$ for the automorphism $x \mapsto h^{-1}xh$ of N, and write the action of such automorphisms exponentially, we see that

$$(h_1 x_1)(h_2 x_2) \;=\; (h_1 h_2)\,(x_1^{\psi(h_2)} x_2). \tag{c2}$$

Conversely, given any groups H and N, and any action of H on N via a homomorphism $\psi\colon H \to \mathrm{Aut}(N)$, we can define a group whose elements are formal products hx $(h \in H,\ x \in N)$, with multiplication defined by (c2). To make the concept of ''formal product'' more precise, let us use the

ordered pair (h, x) instead of the symbol hx. Thus, the law of multiplication becomes

$$(h_1, x_1)(h_2, x_2) \;=\; (h_1 h_2,\; x_1^{\psi(h_2)} x_2). \tag{c3}$$

It is a straightforward exercise to verify that this is indeed a group operation. (Associativity is a quick verification, and the neutral element is, of course, (e_H, e_N). You should find it easy to solve the equation $(h, x)(h', x') = e$ for h' and x', and so get a formula for the inverse of (h, x).) Note that the resulting group has the same underlying set as the direct product group $H \times N$, but has a more general multiplication operation. (''More general'' because, if we take for ψ the trivial homomorphism $\psi\colon H \to \mathrm{Aut}(N)$ which takes all elements of H to the identity automorphism of N, we get precisely the direct product.) It is called ''the *semidirect* product of H and N determined by the map ψ'', and denoted $H \ltimes_\psi N$, or, when there is no danger of ambiguity, $H \ltimes N$.

I should backtrack here, and clarify a point of notation. I said in the very first item of this Companion that functions would in general be written to the *left* of their arguments, and composed accordingly. However, I have made an exception here, and written the action of $\psi(h)$ $(h \in H)$ on elements of N as an exponent, $(\ )^{\psi(h)}$. The reason for this goes back to our decision to write elements of G as products hx with $h \in H$ on the left, and $x \in N$ on the right. The result was that in expressing the product of two such elements as such an element, a member of H had to be ''pulled to the left'' past a member of N, using the formula $xh = h(h^{-1}xh)$; so to each $h \in H$ we associated the automorphism $\psi(h)\colon x \mapsto h^{-1}xh$. Now it is easy to check that if we apply to an element x this automorphism $\psi(h)$, and then apply to the result the automorphism $\psi(h')$ for another element $h' \in H$, this is equivalent to applying to the original element the automorphism $\psi(hh')$. Hence if we used left functional notation, we would have the unpleasant law $\psi(h')\psi(h)(x) = \psi(hh')(x)$; but when we write these automorphisms on the right (and use exponential notation to avoid confusion with multiplication), this takes the nicer form $(x^{\psi(h)})^{\psi(h')} = x^{\psi(hh')}$. So when I called ψ a homomorphism from H to $\mathrm{Aut}(N)$, I should have specified that by $\mathrm{Aut}(N)$ I meant the automorphism group under the composition operation one gets when one writes automorphisms to the *right* of their arguments; the reverse of the ordinary composition. I will note at the end of this discussion some variant forms of the semidirect product construction that one gets by making different notational choices.

We have motivated the general semidirect product construction by considering a group G with a retraction to a subgroup H. Conversely, if we form a semidirect product group $H \ltimes_\psi N$ from any two groups H and N and a homomorphism ψ as described above, we find that the map $(h, x) \mapsto (h, e)$ is a retraction of this group onto a subgroup isomorphic to H, and having kernel isomorphic to N. Hence an isomorphism of a group G with such a semidirect product is *equivalent* to a retraction of G onto a subgroup. As with direct products, we can, when the distinction needs to be made, call a group G constructed from two given groups H and N and a map ψ an ''external'' semidirect product, while when we are given a group G and discover that it has a retraction to a subgroup H, and hence is isomorphic to a group constructed in that way, we can call G an ''internal'' semidirect product.

What are some examples of groups expressible as semidirect products? As noted above, direct products are trivial examples of this construction. A nontrivial example is S_3: the endomorphism which sends every even permutation to the identity, and every odd permutation to (12), is easily seen to be a retraction to the subgroup $\langle(12)\rangle$. That subgroup is isomorphic to Z_2, the kernel of the retraction, A_3, is isomorphic to Z_3, and the action of $\langle(12)\rangle$ on A_3 by conjugation corresponds to the homomorphism $\psi\colon Z_2 \to \mathrm{Aut}(Z_3)$ that sends $[0]$ to the identity automorphism, and $[1]$ to the automorphism $[n] \mapsto [-n]$; hence $S_3 \cong Z_2 \ltimes_\psi Z_3$ for ψ so defined. Recall that S_3 is the symmetry group of an equilateral triangle. More generally, the symmetry group of the regular n-gon can be written as a semidirect product $Z_2 \ltimes_\psi Z_n$, where ψ is defined by the same formula as above. (This group is called the *dihedral group* of order $2n$. Likewise, $Z_2 \ltimes_\psi \mathbf{Z}$, with ψ defined in the analogous way, is called the *infinite dihedral group*.) The group Lang discusses in the top paragraph of p.15 can be seen to be a semidirect product of the additive group of real numbers and the multiplicative group of nonzero real

numbers. Some examples of nonsimple groups that are *not* semidirect products of proper subgroups are the groups Z_{p^n} (p a prime, $n>1$), and the eight-element ''quaternion group'' that Lang introduces at the bottom of p.9.

Since semidirect product decompositions are extremely useful in describing groups, it is worth seeing various formulations of the condition for a group G to be a semidirect product. We have seen that if G can be *retracted* to a subgroup H, then G is a semidirect product of H and the kernel of the retraction. You should verify that an endomorphism r of a group G is a retraction from G to its image $r(G)$ if and only if $r^2 = r$. An entity satisfying the equation $r^2 = r$ is called *idempotent* (Latin for ''equal to its power(s)''); so semidirect product decompositions of G correspond to idempotent endomorphisms of G.

Next, suppose G and H are any two groups, and $f: G \to H$ a surjective homomorphism. It may or may not be possible to find within G a subgroup H_0 that is mapped isomorphically to H by the restriction of f. (For instance, if $G = S_3$, $H = Z_2$ and $f: G \to H$ is the homomorphism taking even permutations to [0] and odd permutations to [1], then $H_0 = \langle(12)\rangle$ is such a subgroup, while if $G = Z_{p^n}$ (p prime, $n>1$), $H = Z_p$, and f is the map sending each congruence class $[i]_{p^n}$ to its image $[i]_p$, there is no such subgroup.) If we have such an H_0, let us denote by g the homomorphism $g: H \to G$ that takes each $h \in H$ to its unique inverse image in H_0 under f. Then we see that

(c4) $$fg = \mathrm{id}_H.$$

What can we say about the other composite, gf? This turns out to be a *retraction* of G onto H_0! In other words, whenever a group homomorphism $f: G \to H$ has a right inverse g, the composite gf is a retraction of G to $g(H) \cong H$; so, in this situation, G is isomorphic to a semidirect product of H with the kernel of this retraction, which is also the kernel of f. (Terminological note: A homomorphism f having a right inverse g as in (c4) is called a *split surjective* homomorphism. Similarly, a homomorphism g having a left inverse f is called *split injective*.)

Finally, given a group G and two subgroups, H and N, we may ask when these will be the image and kernel of a retraction, and thus the ingredients in a semidirect product decomposition of G. You should not find it hard to check that the necessary and sufficient conditions are:

(c5) $$N \text{ is normal in } G, \qquad HN = G, \qquad \text{and} \qquad H \cap N = \{e\},$$

(cf. Lang, Exercise I:L12, (a) and (b)), and that H and N then uniquely determine the retraction.

I noted earlier that the form that our construction of the semidirect product took – in particular, the expression of the action of H on N as a homomorphism into the group of automorphisms of N composed in the way that goes with writing automorphisms on the *right* of their arguments – was a result of certain choices we had made. Different choices lead to different notations, so I should point out some other forms of this construction that you may encounter. To make the distinctions explicit, let us temporarily write $\mathrm{Aut}(N)_r$ for the group of automorphisms of N with multiplication defined in accordance with action of automorphisms on the right, and $\mathrm{Aut}(N)_l$ for the same set but with the multiplication that corresponds to writing automorphisms on the left, the opposite of the former multiplication. Thus, given a group G, a subgroup H, and a normal subgroup N, the map taking each $h \in H$ to the automorphism $x \mapsto h^{-1}xh$ of N is a homomorphism $H \to \mathrm{Aut}(N)_r$, while the map taking h to $x \mapsto hxh^{-1}$ is a homomorphism $H \to \mathrm{Aut}(N)_l$. Let us call these two homomorphisms ψ and θ respectively.

Thus, if we had insisted on writing automorphisms on the left of their arguments, we could have described the law of composition for products hx ($h \in H$, $x \in N$) in a group G with a retraction to a subgroup H as $(h_1x_1)(h_2x_2) = (h_1h_2)(\theta(h_2)^{-1}(x_1)\,x_2)$, and this would have led to an external semidirect product construction in which the law of multiplication (c3) took the form

(c6) $$(h_1, x_1)(h_2, x_2) = (h_1h_2,\ \theta(h_2)^{-1}(x_1)\,x_2)$$

for θ an arbitrary homomorphism $H \to \mathrm{Aut}(N)_l$. Another option would have been to write the typical

element of our group G with a retraction r not as hx but as xh ($x \in N$, $h \in H$). We would then have found that the multiplication of such products takes the form $(x_1 h_1)(x_2 h_2) = (x_1 \theta(h_1)(x_2))(h_1 h_2)$, leading to a construction of the general semidirect product as having underlying set $N \times H$, and multiplication

(c7) $$(x_1, h_1)(x_2, h_2) = (x_1 \theta(h_1)(x_2),\ h_1 h_2).$$

This version of the semidirect product is denoted $N \rtimes_\theta H$. (It is introduced by Lang in Exercise I:L12, but unfortunately, in all versions through the first Springer printing, the formula he gives for the multiplication is garbled.)

The above three constructions are all essentially equivalent. Indeed, if H and N are groups, ψ a homomorphism $H \to \mathrm{Aut}(N)_r$, and θ the homomorphism $H \to \mathrm{Aut}(N)_l$ given by $\theta(h) = \psi(h)^{-1}$, then (c3) and (c6) say exactly the same thing, and the group they define is *isomorphic* to the group $N \rtimes_\theta H$ defined by (c7). (To find the isomorphism between these groups, think about how to turn a product hx into a product $x'h'$ in the situation with which we began this discussion.) As a mnemonic, to remember which side which group goes on when one uses the symbols $\ltimes$ and $\rtimes$, you can think of these as containing normal subgroup symbols $\triangleright$ and $\triangleleft$, with the small end pointing to the subgroup that is to be normal; or, as suggested by a graduate student here, you can think of the open end of either symbol as an open jaw, and remember that ''the beastie eats the kernel''. I should add that these symbols, though common, are not universal. Lang introduces no symbol for the construction in Exercise I:L12; Hungerford uses $N \times_\psi H$; and another writer may just write HN, saying in words that the product is semidirect and has multiplication induced by ψ.

Throughout these notes, semidirect products will be written $H \ltimes_\psi N$ or $H \ltimes N$, and have multiplication defined by (c3).

Re §I.6. Sylow subgroups.

P.33, Proof of Lemma I.6.1 [<]: Until now, Lang has consistently written e for the unit element (also called the identity element) of a group that is not written additively. From this point on, he will vacillate between e and 1. Both notations are common in group theory.

P.33, third line from bottom, ''... is divisible by p'' [>]: The point is that if all elements had orders (''periods'') not divisible by p, then the least common multiple of these orders would be an exponent n of G not divisible by p, hence the order of G would divide a power of n, contradicting the assumption that it is divisible by p.

P.34, middle (end of proof of Theorem I.6.2) [>]:

Alternative proof of the existence Sylow subgroups.

Note that if G is a group, then the set of *subsets* of G can be made a G-set by letting G act by left translation. It is easy to verify:

(c8) If G is a finite group and x any subset of G, then the isotropy subgroup G_x of x under left translation is the largest subgroup H of G such that x is a union of right cosets of H. In particular, the cardinality of x is a multiple of the order of G_x.

We can now give an elegant proof that every finite group G has a p-Sylow subgroup. Let $(G{:}1) = n = p^r s$, where p does not divide s. Let us write $G^{\{p^r\}}$ for the set all subsets of G of cardinality p^r. Then G acts by left translation on $G^{\{p^r\}}$. We shall soon prove that the cardinality of this set, $\binom{p^r s}{p^r}$, is $\equiv s \pmod p$, hence not divisible by p. Assuming this for the moment, we see that some orbit Gx ($x \in G^{\{p^r\}}$) must have cardinality not divisible by p, hence G_x has order a multiple of

p^r. But by (c8), the order of G_x is a divisor of $\#(x) = p^r$, hence it is exactly p^r; that is, G_x is a p-Sylow subgroup of G.

But how do we prove the number-theoretic result $\binom{p^r s}{p^r} \equiv s \pmod{p}$ used above? Surprisingly, we can get this by turning our original proof backwards. We start with some group G of cardinality $p^r s$ that we *know* has a p-Sylow subgroup P (for instance $Z_{p^r} \times Z_s$). If we let P act on $G^{\{p^r\}}$ by left translation, we see that the only orbits of cardinality 1 are the right cosets of P, of which there are exactly s. All other orbits have cardinalities divisible by p; so the cardinality of $G^{\{p^r\}}$, namely $\binom{p^r s}{p^r}$, must be congruent to s mod p.

The additional results on Sylow subgroups that Lang proves in Theorem I.6.4 can also be gotten by looking at the G-set $G^{\{p^r\}}$, but the resulting arguments are fairly close to those Lang gives, so I will not write them out here.

P.36, end of §I.6 [>]: The Sylow theorems are a powerful tool in the study of finite groups. Let us use them to determine the

Structures of groups of order pq.

We know that the groups of prime order are just the abelian groups Z_p (Lang, Proposition I.4.1, last sentence). Exercise I.6:1(a) = I:L24 similarly shows that groups of prime-squared order are just the abelian groups $Z_p \times Z_p$ and Z_{p^2}. We shall show below that groups whose factorizations have the next more complicated form, pq, where p and q are distinct primes, can be described as semidirect products of Z_p and Z_q. We will need the following observation (essentially, Lang's exercise I:L4):

(c9) If G is a finite group with subgroups H_1 and H_2, the set $H_1 H_2$ has cardinality $(H_1 : 1)(H_2 : 1)/(H_1 \cap H_2 : 1)$.

To see this, let us ask what elements fall together under the map $H_1 \times H_2 \to G$ given by $(h_1, h_2) \mapsto h_1 h_2$. If $h_1' h_2' = h_1 h_2$, then we have $h_1^{-1} h_1' = h_2 h_2'^{-1}$, and we see that this element (let us call it k) must belong to $H_1 \cap H_2$. One easily deduces that the elements of $H_1 \times H_2$ that fall together with (h_1, h_2) under the multiplication map are precisely those of the form $(h_1 k, k^{-1} h_2)$ with $k \in H_1 \cap H_2$; so the multiplication map is $(H_1 \cap H_2 : 1)$-to-one, proving (c9). We can now prove

Lemma I.6.c1. *Let G be a finite group, H a subgroup, and N a normal subgroup of G. Then*

(a) *G is a semidirect product $H \leftthreetimes N$ if and only if $(G:1) = (H:1)(N:1)$ and $H \cap N = \{e\}$.*

(b) *In particular, the above conclusion holds if $(G:1) = (H:1)(N:1)$ with $(H:1)$ and $(N:1)$ relatively prime.*

Proof. Since N is assumed normal, the conditions given in (a) are the same as those given in (c5) for G to be a semidirect product, except that the equation $HN = G$ has been replaced by $(G:1) = (H:1)(N:1)$. But from (c9), we see that when $H \cap N = \{e\}$, these two conditions are equivalent.

Assertion (b) is the special case of (a) where the condition $H \cap N = \{e\}$ follows from the fact that the order of this intersection must be a common divisor of the orders of H and N. □

In what situations can we prove that a group G will have subgroups H and N satisfying the above conditions? If the order of G is a product of powers of two distinct primes, p and q, then the corresponding Sylow subgroups of G will have relatively prime orders whose product is the order of G, so we only need to know that one of these subgroups is normal. A convenient tool for showing this is Lemma I.6.7, p.36. Here is an alternative proof of that lemma (part (c) below), with some additional results given by the same method (parts (a) and (b)).

Lemma I.6.c2. *Let G be a group and K a subgroup of finite index in G. Then*

(a) *G has a* normal *subgroup N of finite index which is contained in K, and such that $(G:N)$ divides $(G:K)!$.*

Assuming G itself finite, we have the consequences

(b) *If every prime dividing the order of K is $\geq (G:K)$, then K is normal in G.*

(c) *Hence if $(G:K)$ equals the least prime dividing the order of G, then K is normal.*

Proof. Let $n = (G:K)$, and consider the action of G on the n-element G-set G/K. This is a homomorphism *into* the permutation group on that G-set, which has order $n!$, hence *onto* a subgroup S of that permutation group. The order of S will be a divisor of $n!$, and the kernel of that homomorphism will be a normal subgroup N whose index in G is that order, giving (a).

Note that $(K:N) = (G:N)/(G:K) = (G:N)/n$, a divisor of $(n-1)!$, hence a product of primes $\leq n-1$. But in the situation of statement (b), every prime divisor of $(K:N)$ must be $> n-1$, so there are no such divisors, and $K = N$. Statement (c) is a particular case of (b). □

We can now characterize all groups whose orders are products of two distinct prime factors.

Lemma I.6.c3. *Let $p < q$ be primes. Then every group of order pq is a semidirect product of a cyclic subgroup of order p and a normal cyclic subgroup of order q. If $q \not\equiv 1 \pmod{p}$, the only such group is the* direct *product of these subgroups, an abelian group isomorphic to $\mathbf{Z}/pq\mathbf{Z}$. If $q \equiv 1 \pmod{p}$, there also exists a nonabelian group of order pq.*

Proof. Let G be a group of order pq, and let H be a p-Sylow subgroup and N a q-Sylow subgroup of G. By Lemma I.6.c2(c) above, N is normal, hence by Lemma I.6.c1(b), G is a semidirect product of the indicated form.

Hence let us write $G \cong H \ltimes_{\psi} N$ where ψ is a homomorphism $H \to \mathrm{Aut}(N)$. This group will be nonabelian if and only if ψ is nontrivial, and such a nontrivial homomorphism can be found if and only if $\mathrm{Aut}(N)$ has order divisible by p (Lang, Theorem I.6.2 and Corollary I.6.6). By Proposition I.4.3(iii), the automorphism group of the cyclic group of order q has the same number of elements as the group has generators; since q is prime, all nonidentity elements are generators, so $\mathrm{Aut}(N)$ has order $q-1$. This will be divisible by p if and only if $q \equiv 1 \pmod{p}$, establishing the asserted condition for the existence of a nonabelian group.

On the other hand, if G is abelian, and hence is the direct product of H and N, we know from Proposition I.4.3(v) that G is cyclic. □

We now ask: If $q \equiv 1 \pmod{p}$, will there be just one, or more than one isomorphism class of nonabelian groups of order pq? To approach this question, suppose G is a nonabelian group of order pq, with a normal subgroup N cyclic of order q, and a p-Sylow subgroup H cyclic of order p. Let a be a generator of N. Any automorphism φ of N carries a to another generator of N, which will have the form a^r for some r relatively prime to q; given φ and a, this r is clearly uniquely determined mod q. Note that φ will carry any other generator $a' = a^i$ of N to $(a^r)^i = a^{ir} = a'^r$; hence the residue of r modulo q is in fact an invariant of the automorphism φ, independent of our choice of generator a of N. We see that composing two automorphisms corresponds to multiplying the corresponding residues modulo q. Now if we map each element hn of $G = HN$ to the automorphism of N that it induces, we can see that this will depend only on h (since N is abelian), and will determine by the preceding results a homomorphism of G into the group of nonzero residues modulo q. The image of this homomorphism will be a subgroup of this group of residues; since it equals the image of H under the same map, it will have order $(H:1) = p$.

This subgroup of the multiplicative group of integers modulo q will be an invariant of the structure of the group G; i.e., if G and G' are each a semidirect product of a cyclic subgroup of order p with a

cyclic subgroup of order q, and they are isomorphic, they must determine the same order-p group of residues modulo q.

The converse is also true: if G and G' are semidirect products of a cyclic group of order p and a cyclic group of order q, and they determine the same group of residues, then they will be isomorphic. Indeed, let a be a generator of a normal subgroup $N \lhd G$ of order q, and b a generator of a subgroup $H < G$ of order p, and say conjugation by b carries a to a^r. We now want to find elements a', b' in our other group G' that will correspond to a and b under an isomorphism. Let a' be any generator of a normal subgroup $N' \lhd G'$ of order q, but within a subgroup $H' < G'$ of order p, let us be more careful: Because of our assumption about G and G' determining the same groups of residues, we know that there will be some element of H' conjugation by which carries a' to a'^r; let b' be such an element. Then it is not hard to verify that the map $G \to G'$ carrying each element $b^i a^j$ to $b'^i a'^j$ is an isomorphism.

The problem of whether there can be *more than one* isomorphism class of nonabelian groups of order pq therefore reduces to the question: Does the multiplicative group of nonzero residues modulo q contain more than one subgroup of order p?

Since this group of residues is abelian, it has a unique p-Sylow subgroup, which will contain all subgroups of order p. So the problem is what this p-Sylow subgroup looks like! If it has order p (i.e., if $q-1$, though divisible by p, is not divisible by p^2), then clearly it will be the unique subgroup of order p, and thus by the above reasoning there will be a unique isomorphism class of nonabelian groups of order pq. For instance, this happens if pq = 6, 14, 21 or 39. (E.g., it happens in the last case because $39 = 3 \cdot 13$, and 3 divides $13-1$ just once.) What about cases where $q-1$ *is* divisible by p^2, e.g., pq = 10, 26 or 57? Checking by hand, we find that in each of the cases just listed, the p-Sylow subgroup of $\mathrm{Aut}(\mathbf{Z}/q\mathbf{Z})$ has only one subgroup of order p. In our notes on §IV.1, we shall see that simple ring-theoretic considerations prove that this is always so, and hence that under the conditions of the last sentence of Lemma I.6.c3, there is, up to isomorphism, a *unique* nonabelian group of order pq. In the mean time, we will live with the suspense of wondering how that will be proved!

The above results, together with Exercises I.6:1-2, give us the structures of groups of most ''small'' orders. The first case not covered, order 12, is not hard to analyze; I like to work it out in class. The next value, 16, is the fourth power of a prime, a difficult case, which is more work than we have time for, but the two after that, 18 and 20, can be analyzed with the help of some further easy results, which I give below, but do not make required reading for Math 250A.

Further notes on structures of finite groups. (Optional reading for my 250A.)

Lemma I.6.c4. *Let p and q be primes and i, j positive integers, and let G be a group of order $p^i q^j$. Then G has a normal q-Sylow subgroup if* either (a) $i = 1$ *and* $p < q$, or (b) $i = 2$ *and* $p < q-1$.

Thus, if either of these conditions holds, G is a semidirect product of this normal q-Sylow subgroup and any p-Sylow subgroup of G.

Proof. Case (a) can be done by the same method as Lemma I.6.c3 above, but let us indicate a different method, which parallels the one we shall use for (b). We know from Theorem I.6.4(iii) (Lang, p.35) that the number of conjugates of a q-Sylow subgroup Q of G is $\equiv 1 \pmod q$. But this number is the index in G of the normalizer of Q, which is a divisor of the index of Q itself, which is p. The only divisors of p are p and 1, and the former cannot be congruent to 1 (mod q) because it is strictly between 1 and q, hence the number of conjugates of Q must be 1; i.e., Q must be normal.

In case (b) we begin in the same way by noting that the number of conjugates of Q must be p^2, p or 1. As before, we cannot have $p \equiv 1 \pmod q$. Can p^2 be $\equiv 1 \pmod q$? This would mean that q divided $p^2 - 1 = (p-1)(p+1)$. Hence one of these factors must be divisible by q, hence must be *at least* q. But our hypothesis $p < q-1$ says that even the larger of these two factors, $p+1$, is $< q$. So as in

case (a), Q must be normal.

The final sentence follows by Lemma I.6.c2. □

Corollary I.6.c5. *Any group of order p^2q, where p and q are primes, has a normal p- or q-Sylow subgroup, and is thus a semidirect product of an abelian p-group and an abelian q-group.*

Sketch of Proof. If $p = q$, the first assertion is trivial, while the second follows from Exercise I.6:2. If $p > q$, then by *interchanging* the roles of p and q we are reduced to case (a) of the preceding lemma. If $p < q$, then we are OK by part (b) of that lemma unless $p = q-1$. But if two primes differ by 1, one of them must be even, so they must be 2 and 3 respectively. Thus, the order of G is $2^2 \cdot 3 = 12$, a case which, as I mentioned earlier, I like to analyze in class. (In that case, one finds examples both with 2-Sylow subgroup normal and with 3-Sylow subgroup normal.) □

I will not discuss here the question of *how many* nonisomorphic groups of order p^2q there are; in other words, the classification of actions of abelian groups of order p^2 on abelian groups of order q, and vice versa.

One might hope to stretch the method of Lemma I.6.c4 further, e.g., to prove that (perhaps with a small number of exceptions), for $p < q$, groups of order p^3q^j always have normal q-Sylow subgroups. But the argument does not generalize, essentially because in the factorization of $p^3 - 1$, namely $(p-1)(p^2 + p + 1)$, the second factor may be divisible by a prime significantly larger than p. In fact, whenever q is a prime dividing $p^2 + p + 1$, one can show that $\mathrm{Aut}((Z_p)^3)$ has order divisible by q, hence one gets a nontrivial map from Z_q to this automorphism group, which allows one to construct a group $Z_q \ltimes (Z_p)^3$ of order p^3q with non-normal q-Sylow subgroup.

Of course, in this situation one may try to prove that every group of order p^3q is *either* such a semidirect product $Z_q \ltimes (Z_p)^3$ *or* of the form $H \ltimes N$ where N is a normal q-Sylow subgroup. We shall note in the next paragraph one example for which this does not hold; I don't know whether it holds in all or most other cases. One can also give conditions which guarantee that $p^2 + p + 1$ is not divisible by q. One such condition is clearly $q > p^2 + p + 1$. Another is $q \equiv 2 \pmod 3$; you might try to discover a proof of this.

The smallest group that has *no* normal Sylow subgroup is S_4. Indeed, a 3-Sylow subgroup of S_4 is given by cyclic permutation of three of the four elements on which S_4 acts, and will not be normalized by a permutation that sends one of these three elements to the remaining element. Likewise a 2-Sylow subgroup will be the symmetry group of a square, where the elements $\{1, 2, 3, 4\}$ are arranged as the vertices of that square, and it is easy to see that a permutation that does not preserve the pairing of opposite vertices will not normalize it.

However, S_4 has a normal subgroup of index 2, which in turn has a normal 2-Sylow subgroup; thus S_4 is solvable. The smallest nonsolvable group is the 60-element group A_5.

Let us end by noting two further useful consequences of Lemma I.6.c2. The first is the simplest possible case of that lemma, and is a standard elementary exercise. (It is worked as an Example on p.28 of Lang.) The second yields a large class of examples that show that groups G of order n need not have subgroups of all orders dividing n.

Corollary I.6.c6 (to Lemma I.6.c2)**.** (a) *Any subgroup of index* 2 *in a group is normal.*

(b) *If a* simple *group G has a subgroup of finite index r, then* $(G:1) \mid r!$. □

In particular, we see from (b) that the simple group A_5 of order 60 can have no subgroups of order 15 or 20.

It is an amusing exercise to discover the structures of Sylow subgroups of symmetric groups. I like to

do this interactively with the class, so I will not present it here; but I will give a terse summary of the result, which you can read after seeing the development.

Given any two finite groups G and H, let us form the direct product H^G of a family of copies of H indexed by the elements of G. Let G act on this product by translating the subscripts; i.e., by letting $g \in G$ send each G-tuple $(h_\gamma)_{\gamma \in G}$ to $(h_{g\gamma})_{\gamma \in G}$. Since this is an action of G on H^G by automorphisms, we can form the semidirect product induced by this action; this is called the *wreath product*, G wr H. (That term is reasonable in the light of the pictures we drew!)

Now let us write $P_{n,p}$ for the p-Sylow subgroup of S_n. One finds that for n a power of p, these groups are described inductively by

$$P_{p^m,p} \cong Z_p \text{ wr } P_{p^{m-1},p}.$$

To get a description for general n, we write n to base p: $n = a_m p^m + \ldots + a_0$, with $0 \leq a_i < p$. We then have

$$P_{n,p} \cong \prod_i (P_{p^i,p})^{a_i}.$$

*Note on the much used word ''**structure**''.* In the last few pages, we have used that word many times. Let us observe that it has several senses in mathematics. On the one hand, looking at a group as a set with a multiplication operation satisfying appropriate conditions, one can call that multiplication operation a ''group structure'' on the set. (Similarly, a distance function gives a set a ''structure of metric space'', etc. etc..) On the other hand, when we spoke of studying the ''structures of groups of small orders'', this meant finding convenient descriptions of all such groups, up to isomorphism. (So in this usage, does ''structures'' just mean ''distinct isomorphism classes''? Not exclusively; for instance the above investigation of the Sylow subgroups of S_n helps us understand the ''structure'' of S_n, even though that involves no classification question.) In still another usage, one can speak of groups, fields, metric spaces, manifolds, etc. as various kinds of ''mathematical structures''. And these three senses merge into one another at times.

I consider it valuable to have words like this that do not have precise technical definitions, and so allow us to talk nontechnically about mathematics. In fact, I think it is unfortunate when someone takes one of these words and gives it a technical definition in some field of mathematics, depriving our communication of some of its flexibility.

Re §I.7. Direct sums and free abelian groups.

Before reading §I.7, I recommend reading Lang's set-theoretic Appendices 2.2-2.4, together with the comments on them in this Companion. (Those comments, in turn, indicate a point at which you should switch from Lang to my handout ''The Axiom of Choice, Zorn's Lemma, and all that''.) §I.7 can be read without that set-theoretic material, but then the important result Theorem I.7.3 is only proved in the finitely generated case. The discussion of that theorem below shows how to extend the proof to the general case.

P.38, line 4 [~]: Delete the condition ''if the family is not empty''! The empty set *is* a basis for the trivial group. (Lang will impose similar inappropriate nonemptiness conditions in several definitions in later sections as well. I will postpone discussing *why* I consider them inappropriate till later, though you can read that discussion, at the end of the material on §III.5, now if you wish.)

P.38, middle [~]: What Lang denotes $\mathbf{Z}\langle S \rangle$ I recommend writing as $\mathbf{Z}\,S$, to avoid conflict with ring-theorists' notation for free rings, and to agree with standard notation for group rings. (In any case, he uses the notation $\mathbf{Z}\langle S \rangle$ only briefly.)

P.38, last paragraph [>]: The comment that begins with the third sentence of this paragraph, ''If $g\colon S \to B$ is a mapping ...'', is very important! What is shown is that any set-map from a basis of a free abelian group F into any abelian group B extends uniquely to a *homomorphism* $F \to B$; intuitively, that we can choose homomorphisms $F \to B$ that send the elements of S to ''wherever we like''. It should be stated as a theorem; it will be used both by Lang and in this Companion. The result is known as

the *universal property* of free abelian groups, and I will refer to it by that name below.

P.40, *Proof* of two italicized lines [~]: You have probably seen the construction of the field of fractions of a commutative integral domain (or at least the special case of the construction of $\mathbf{Q}$ from $\mathbf{Z}$) as a set of equivalence classes of ordered pairs. The construction Lang is about to describe is analogous to this; in fact, when applied to the abelian monoid of nonzero elements of such an integral domain under multiplication, it is exactly that construction (minus the consideration of 0 and addition). When applied to the *additive* monoid of nonnegative integers, it likewise gives the additive monoid of all integers; i.e., it ''brings in'' the negative integers. Lang assumes in the first sentence that students are more likely to have seen the integers-from-nonnegative integers construction than the field-of-fraction construction; though I think that under current US undergraduate curricula, the reverse is true.

In the second sentence, where Lang defines what is meant by ''(x,y) is equivalent to (x',y')'', you should check that this is indeed an equivalence relation. The verification uses the assumption that the cancellation law holds in M.

P.40, end of discussion of the Grothendieck construction [>]: Here are a few examples of that construction; you should be able to work out the verifications: If $M = \mathbf{N}$, i.e., the monoid of nonnegative integers, then $K(M) = \mathbf{Z}$. If M is a monoid generated by an element x such that the powers of x repeat cyclically with period n, starting with x^r (i.e., such that $x^r = x^{r+n}$, and thus $x^{r+i} = x^{r+i+n}$ for all $i \geq 0$, but no x^i with $j < r$ is equal to any other power of x), then $K(M)$ is the cyclic group of order n, i.e., the equation $x^j = x^{j+n}$ holds in $K(M)$ for all j. In particular, if $n = 1$, i.e., if the powers of x become constant from some point on, then $K(M)$ is the trivial group. If M is in fact a group, then $K(M) \cong M$; that is, the canonical map is an isomorphism.

P.41, proof of Lemma I.7.2 [~]: Recall from our discussion of semidirect products that a group homomorphism $f: G \to H$ is called a *split surjection* if f has a right inverse, $g: H \to G$, and that when this happens, gf is a *retraction* of G onto $g(H)$, and G is a semidirect product of $g(H)$ and the kernel N of f.

Now if G is abelian, this semidirect product will be the direct product (equivalently, the direct sum), since the action of H on G, and hence on N, by conjugation is trivial. Thus, a splitting of a surjective homomorphism of abelian groups, $f: A \to B$ corresponds to a decomposition of A as the direct sum of $C = \mathrm{Ker}(f)$ and a subgroup that is isomorphic to B (and, indeed, is mapped isomorphically via f to B).

Now suppose $f: A \to B$ is a surjective homomorphism of abelian groups, and B is *free*, say with a basis E. By surjectivity of f, we can find a set-map $E \to A$ taking each element of E to an inverse image in A, and by the universal property of free abelian groups, this extends to a homomorphism $g: B \to A$. Since $fg : B \to B$ agrees with the identity endomorphism of B on the basis E, it must equal the identity, so g is a right inverse to f. Thus, by our preceding observations, we get a decomposition of A as the direct sum of a subgroup isomorphic to B, and the subgroup $\mathrm{Ker}(f)$. This is the essentially Lang's proof of Lemma I.7.2, but he writes out as computations what we have embodied in general observations.

Remarks on the concepts of split surjections, split injections, and split exact sequences.
(Footnote to the above discussion.)

We have been using the terms ''split surjection'' for a homomorphism with a right inverse, and ''split injection'' for a homomorphism with a left inverse. It is clear that these are indeed surjections and injections; but why the term ''split''?

Recall that the relation between an abelian group A, its image B under a surjective homomorphism, and the kernel C of that homomorphism can be expressed by the *exact sequence*

$$0 \to C \to A \to B \to 0. \tag{c10}$$

Exact sequences beginning and ending with ''0'' and with three groups in between are called ''short exact sequences''. Note that given any one-to-one group homomorphism f, i.e., any exact sequence $0 \to C \to$

A, if the image $f(C)$ is normal in A (as is automatic for abelian groups), then one can embed this in a short exact sequence (c10) in an essentially unique way, namely by taking $B = A/f(C)$. Likewise, given an onto homomorphism f, i.e., an exact sequence $A \to B \to 0$, one can embed this (without any additional assumptions) in a short exact sequence (c10) by taking $C = \mathrm{Ker}(f)$.

A short exact sequence (c10) is said to *split* if it is isomorphic to a direct sum of two ''trivial'' short exact sequences, $0 \to C \to C \to 0 \to 0$ and $0 \to 0 \to B \to B \to 0$ (with the nontrivial map the identity in each case), i.e., if it is isomorphic to a sequence

$$0 \to C \to C \oplus B \to B \to 0, \tag{c11}$$

where the first map has the form $c \mapsto (c, 0)$ and the second the form $(c, b) \mapsto b$.

This use of the word ''split'' is, hopefully, quite natural. We can now justify the terms ''split surjective homomorphism'' and ''split injective homomorphism'' by showing that a surjective or injective homomorphism has an inverse on the appropriate side if and only if the corresponding short exact sequence splits. Indeed, it is straightforward to verify

Lemma I.7.c1. *Let*

$$0 \to C \xrightarrow{f} A \xrightarrow{f'} B \to 0 \tag{c12}$$

be a short exact sequence of abelian groups. Then the following conditions are equivalent.

(a) *The exact sequence* (c12) *splits; i.e., up to isomorphism, it has the form* (c11).

(b) *The image of f, equivalently, the kernel of f', is a direct summand in A.*

(c) *The surjective map f' splits, i.e., has a right inverse as a group homomorphism.*

(d) *The injective map f splits, i.e., has a left inverse as a group homomorphism.* □

P.41, proof of Theorem I.7.3 (continued on p.42) [=]: Let us do this in general, rather than limiting ourselves to the finitely generated case. Let X be a basis of A, and let us choose a *well-ordering* $\leq$ of X (cf. the handout on the Axiom of Choice, and/or Lang's Appendix 2.4). For each $s \in X$, let

$$A_{<s} = \text{subgroup of } A \text{ generated by } \{x \in X \mid x < s\},$$
$$A_{\leq s} = \text{subgroup of } A \text{ generated by } \{x \in X \mid x \leq s\}.$$

Note that $A_{<s} = \cup_{t<s} A_{\leq t}$. Given any $s \in X$, let $f_s : A \to \mathbf{Z}$ be the map taking each $a \in A$ to the *coefficient* of s in the unique expression of a as a linear combination of the elements of X. Thus, the restriction of f_s to $A_{\leq s}$ has kernel $A_{<s}$. Now let

$$S = \{s \in X \mid (B \cap A_{<s}) \neq (B \cap A_{\leq s})\}.$$

Thus, for $s \in S$, the image $f_s(B \cap A_{\leq s}) \subseteq \mathbf{Z}$ is a nonzero subgroup of an infinite cyclic group, hence is infinite cyclic; say $f_s(B \cap A_{\leq s}) = n_s\mathbf{Z}$. For each $s \in S$, let us choose $y_s \in B \cap A_{\leq s}$ such that $f_s(y_s) = n_s$, and let $Y = \{y_s \mid s \in S\}$. We claim that Y is a basis for B; this will prove B free, as required.

To show Y linearly independent, suppose

$$\Sigma_{s \in S}\, m_s y_s = 0 \qquad (m_i \in \mathbf{Z}).$$

If not all m_s are zero, let t be the *largest* element of S (under our ordering of X) such that $m_t \neq 0$. (A largest value exists because only finitely many m_s are nonzero.) Applying f_t to the above equation, we get $m_t n_t = 0$, a contradiction since both m_t and n_t were assumed nonzero.

To show that Y generates B, let us, for every nonzero $a \in A$, define $d(a)$ to be the largest element of X occurring with nonzero coefficient in the expression for a. Note that if $a \in B$, $d(a)$ is the least x such that $a \in A_{\leq x}$; thus $a \notin A_{<d(a)}$, so $d(a) \in S$. Now if Y does not generate B, let b be an element of B not in the subgroup generated by Y, and chosen, among all such elements, to *minimize* $d(b)$. (This is where we use the fact that $\leq$ is a *well*-ordering!) Thus $f_{d(b)}(b)$ has the form $r\, n_{d(b)}$. Hence

$b - ry_{d(b)}$ lies in $A_{<d(b)}$, i.e., $d(b - ry_{d(b)}) < d(b)$. Thus, by choice of b, the element $b - ry_{d(b)}$ *does* belong to the subgroup generated by Y, and adding $ry_{d(b)}$, we conclude that b lies in that subgroup; a contradiction. This completes the proof that B is free, with basis Y.

We now come to the proof of uniqueness of the rank (cardinality of a basis), which Lang gives in the top paragraph of p.42. I am torn between saying that Lang's proof, based on looking at the cardinality of B/pB is ''instructive'' and that it is ''a cheap trick''. If we regard the study of free abelian groups as the $R = \mathbf{Z}$ case of the study of free modules over a ring (to be defined in Chapter III), this counting trick cannot be extended to the general case; it just happens to work in this case because the ring $\mathbf{Z}$ has homomorphic images $\mathbf{Z}/p\mathbf{Z}$ which are finite. On the other hand, the idea of finding invariants of one kind of mathematical structure by applying to them a construction (what we will soon call a *functor*) taking them to structures of another sort that we have better information about, is a tool of great generality and power.

Well, let us be satisfied with this proof for now; we will see a different one that works for modules over general commutative rings later. Note, however, that this proof works only if one of the two bases is finite. This is not a fault of this particular proof; it is true of every proof of the finite case that I know, including those used in module-theoretic contexts. To get the infinite case, we shall use the following nice observation.

Lemma I.7.c2. *Let G be a group* (*not necessarily abelian*) *generated by an infinite set X. Then if Y is any other generating set for G, there is a subset $Y' \subseteq Y$ of cardinality $\leq \mathrm{card}(X)$ which also generates G.*

(*Likewise, if G is generated by a finite set X, then any generating set Y has a* finite *generating subset, though not necessarily of cardinality $\leq \mathrm{card}(X)$*).

Proof. For each $x \in X$, since x lies in the group generated by Y, it can be written as an expression in finitely many elements of Y; hence there is a finite subset $Y_x \subseteq Y$ such that x lies in the subgroup generated by Y_x. If we take the union of these sets Y_x, over all $x \in X$, we get a set $Y' \subseteq Y$ whose cardinality is $\leq \mathrm{card}(X)$ if X is infinite (cf. Theorem A2.3.3), or finite if X is finite, and such that the subgroup generated by Y' contains X, hence equals G. □

Now if S and T are two infinite bases of a free abelian group B, then the above lemma shows that T has a subset T' of cardinality $\leq \mathrm{card}(S)$ which generates A. But it is easy to verify that a *proper* subset of a basis of A cannot generate A, hence T' must be T, showing that $\mathrm{card}(T) \leq \mathrm{card}(S)$. Similarly, $\mathrm{card}(S) \leq \mathrm{card}(T)$, proving the equality of these cardinalities. □

This completes the proof of Theorem I.7.3.

It is interesting to note how the proof of the first part of that theorem, on existence of bases, which we carried out using Well-Ordering, would be proved using Zorn's Lemma. Given A, X and B as in that proof, let us, for any subset $S \subseteq X$, define A_S to be the subgroup of A generated by S. Let P denote the set of ordered pairs (S, Y) such that $S \subseteq X$, and Y is a basis of $B \cap A_S$, and let us write $(S, Y) \leq (S', Y')$ if $S \subseteq S'$ and $Y \subseteq Y'$. Since $(\emptyset, \emptyset) \in P$, P is nonempty. Clearly any chain $\{(S_i, Y_i) \mid i \in I\}$ in P has an upper bound, given by $\{(\cup S_i, \cup Y_i)\}$. Hence P has a maximal element (S, Y).

Having made this application of Zorn's Lemma, we are not at the end of the proof, but at the beginning of the group-theoretic part! We claim that from the maximality of (S, Y) in P, it follows that $S = X$. For suppose $x \in X - S$. There are two possibilities: If $B \cap A_{S \cup \{x\}} = B \cap A_S$, then we find that $(S \cup \{x\}, Y)$ is an element of P greater than (S, Y), contradicting the maximality of this pair. On the other hand, if $B \cap A_{S \cup \{x\}}$ is strictly larger than $B \cap A_S$, then we look at the nonzero map $f_x : B \cap A_{S \cup \{x\}} \to \mathbf{Z}$, choose an element $y \in B \cap A_{S \cup \{x\}}$ that is mapped to a generator of the cyclic image group, and show that $Y \cup \{y\}$ is a basis of $B \cap A_{S \cup \{x\}}$; so in this case, $(S \cup \{x\}, Y \cup \{b\})$ gives a contradiction to the maximality of (S, Y). (This step is a little easier than proving that Y is a basis for

B in the proof using Well-Ordering, because it simply involves adding one element to a set known to be a basis of a subgroup. You should work out the details.) Thus, the maximal pair (S, Y) satisfies $S = X$, so Y is a basis of $B \cap A_X = B$, as desired.

Re §I.8. Finitely generated abelian groups.

P.43, proof of Theorem I.8.2 (through p.45) [~]: To motivate the result, it is instructive to try to construct various sorts of finite abelian p-groups – e.g., by starting with cyclic p-groups, forming direct sums of these, and then forming factor-groups. We find that all the examples we can come up with can themselves be written (after a change of generators) as direct sums of cyclic p-groups. So we would like to see whether we can prove that every finite abelian p-group A has this form.

To get such a proof, we need a construction that, starting with an arbitrary finite abelian p-group, tells us how to find generators for a family of cyclic direct summands. Such a system of summands is not unique (although having read the statement of Theorem I.8.2, we know that it will be unique up to isomorphism). E.g., if $A = Z_{p^3} \oplus Z_p$, we find that along with the given cyclic summands, generated by $(1, 0)$ and $(0, 1)$ there are others, such as those generated by $(1, 1)$ and $(p^2, 1)$.

However, we cannot start with an arbitrary cyclic subgroup and expect it to form part of a direct sum decomposition. E.g., if $A = Z_{p^2}$, then the cyclic subgroup generated by p is not a direct summand. Looking at what goes wrong in that case, we might conjecture that any *maximal* cyclic subgroup of a finite abelian p-group A would be a direct summand. But even this is not so: in $A = Z_{p^3} \oplus Z_p$, the cyclic subgroup generated by $(p, 1)$ is maximal (since that element is not of the form pa, as we see by looking at its second coordinate), but it cannot be part of a direct-sum decomposition, since it has order p^2, and we know from the statement of Theorem I.8.2 that in any decomposition of A as a direct sum of cyclic subgroups, the summands must have orders p^3 and p.

In the proof of Theorem I.8.2 in Lang, the trick used is to choose a cyclic subgroup of *maximal order*, i.e., one whose order is an exponent of the whole group A. Using this same idea, we shall prove below a stronger result than that theorem, namely that any *not necessarily finite* abelian group of exponent p^n is a direct sum of cyclic subgroups. The key step, which shows that we can ''split away'' arbitrary direct sums of cyclic subgroups of maximal order, is the following:

Lemma I.8.c1. *Let p be a prime and n a positive integer.*

(a) *If B is a direct sum* (*or a direct product*) *of cyclic groups of order p^n, then it satisfies*

(c13) *Every element $x \in B$ of exponent p^i can be written $p^{n-i}y$ for some $y \in B$.*

(b) *If A is an abelian group of exponent p^n, and B a subgroup satisfying* (c13)*, then B is a direct summand in A* (*i.e., there exists a subgroup $C \subseteq A$ such that $A = B \oplus C$*).

Proof. (a) is straightforward. To prove (b), we shall show how to ''build up'' a complementary summand C. Suppose we have a ''partial'' complementary summand; that is, a subgroup $D \subseteq A$ satisfying $B \cap D = \{0\}$, but such that $B + D$ is not all of A. Thus there exist elements of A not in $B + D$. If we multiply such an element by p enough times, we will get zero (since A has exponent p^n); in particular, after a certain (possibly smaller) number of multiplications by p, we get an element of $B + D$. Hence if we stop one step short of this, we get an element $x \notin B + D$ such that $px \in B + D$. Thus

$$px = b + d \quad \text{where} \quad b \in B,\ d \in D.$$

Since x is annihilated by p^n, if we multiply both sides of this equation by p^{n-1} we get $0 = p^{n-1}b + p^{n-1}d$. Hence $p^{n-1}b = -p^{n-1}d \in B \cap D = \{0\}$, so b is annihilated by p^{n-1}. Hence by assumption (c13), we can write $b = pb'$ $(b' \in B)$. Thus, if we let $y = x - b'$; the above displayed equation gives $py = d$. We claim that if we set $D' = D + \langle y \rangle$, then this will be a larger subgroup than

D, but will still satisfy $B \cap D' = \{0\}$.

To see that D' is strictly larger than D, it suffices to note that $B + D'$ contains $b' + y = x$, which by assumption is *not* in $B + D$. This observation can be translated to say that the image $\bar{y}$ of y in $A/(B + D)$ equals the image of x therein, and so is nonzero. Since by assumption $px \in B + D$, this element $\bar{y}$ has order p. We can now prove that $B \cap D' = \{0\}$. Indeed, suppose that an element of D', say $e + ry$ ($e \in D$, $r \in \mathbf{Z}$) lies in B. Then its image in $A/(B + D)$ is 0; but we can also write this image as $r\bar{y}$ (because $e \in D$ has image 0); so $r\bar{y} = 0$. Since $\bar{y}$ has order p, this says that r is divisible by p, hence ry is a multiple of $py \in D$. Hence $e + ry \in D$, hence, as $B \cap D = \{0\}$, we have $e + ry = 0$. Since $e + ry$ was an arbitrary element of $B \cap D'$, this shows that $B \cap D' = \{0\}$, as claimed.

This passage, from a partial complement D of B to a larger partial complement D', gives the general ''step'' of our desired construction. We now put such steps together using Zorn's Lemma: Let P be the set of all subgroups $D \subseteq A$ with the property $B \cap D = \{0\}$, partially ordered by inclusion. It is easy to verify that the union of any chain in P lies in P, and hence forms an upper bound for that chain. So by Zorn's Lemma, P has a maximal element C. We claim that

$$B + C = A.$$

For if this were not so, our preceding argument would give us a larger $C' \in P$, contradicting the maximality of C. Since $B \cap C = \{0\}$ by definition of P, we have $A = B \oplus C$, as desired. □

From this, we can now prove

Theorem I.8.c2. *If A is an abelian group of exponent p^n, then A is a direct sum of cyclic subgroups of orders dividing p^n.*

Proof. Let us begin by constructing the family of subgroups of order exactly p^n that will occur in our decomposition. To do this, let P be the set of all sets X such that X is a set of cyclic subgroups of A, each of order p^n, whose sum in A is direct. It is easy to verify that the union of any chain in P is a member of P, and is thus an upper bound for that chain. Hence P has a maximal element X; let $B \subseteq A$ be the sum of the subgroups comprising X. We note that by definition of P, this is a direct sum. By part (a) of the preceding lemma, B has property (c13), hence by part (b) of that lemma, we can write $A = B \oplus C$. Being a subgroup of A, C has exponent p^n; but we claim it has no element of order exactly p^n. For if c were such an element, then $X \cup \{\langle c \rangle\}$ would be a member of P larger than X, contradicting maximality. It follows that C has exponent p^{n-1}.

We may now assume by induction on n that C is a direct sum of various cyclic subgroups of orders p^i with $i \leq n-1$. Putting together this decomposition of C and our given direct sum decomposition of B, we get the desired decomposition of A. □

We noted earlier that a maximal cyclic subgroup of a group of exponent p^n need not be a direct summand; but our proof of the above theorem (via the preceding lemma) shows that a maximal cyclic subgroup *having largest possible order* is indeed a direct summand. Curiously, if among maximal cyclic subgroups we choose one having *smallest* order, it will also be a direct summand. This leads to another proof of the above theorem, sketched in Exercise I.8:7. We thus see that our example of a maximal cyclic subgroup which was *not* a direct summand had to avoid both these extremes: it in fact had order p^2 in a group whose largest- and smallest-order cyclic summands had orders p^3 and p respectively.

Still another way of proving the above theorem is developed in Exercise I.8:6. (The idea there is to choose a generating set of the exponent-p abelian group $A_p = \{a \in A \mid pa = 0\}$, in a way which takes account of the chain of subgroups consisting of elements that are ''multiples of p^{n-1},'', ''multiples of p^{n-2},'', etc.. One can also give a proof ''dual'' to this, in which one chooses a basis of the exponent-p abelian group A/pA in a way which takes account of the chain of subgroups given by those elements

which are images of elements of A of exponent p, of exponent p^2, etc..)

For the proof of *uniqueness up to isomorphism* of these decompositions, when the group A is finitely generated, see Lang, p.44 (bottom) to 45. The idea is that, if we write A_p for the subgroup $\{a \in A \mid pa = 0\}$, then the number of cyclic summands of each order can be read off from the sizes of the subgroups $A_p \cap p^{i-1}A$. This idea can be adapted to our non-finitely-generated context with a little work; you will not be responsible for this generalization, but if you are interested, you might think about how this would be done.

Of course, from the above theorem we can get, using Theorem I.8.1 (p.42) the

Corollary I.8.c3. *If A is an abelian group of exponent m, then A is a direct sum of cyclic subgroups of orders dividing m.* □

Every time I teach this section, I end up getting fascinated by questions of ways that the main results of the section can and cannot be generalized; the number of exercises relating to this material that I have accumulated is testimony to this fact. There is, in fact, a well-developed theory of abelian groups, which I have never had time to study. If you would like to look further into it, a standard work in the field is L. Fuchs, *Abelian groups*, Pergamon Press, 1960. (A reference to a book that studies similar questions in a module-theoretic context is given at the end of my comments on §III.7 below.)

P.46, line 2 [=]: For ''a finite set'' read ''the finite set S''.

P.46, line preceding Theorem I.8.5 [~]: Lang writes ''mA is free, hence A is free''. The argument is that multiplication by m is a homomorphism $A \to mA$; it is clearly onto, and it is one-to-one because A is torsion-free. Hence $A \cong mA$, so if the latter is free, the former is, also.

Re §I.9. The dual group.

P.46, beginning of the section [~]: On pp.24-25, the ''standard'' cyclic group of order n was written $\mathbf{Z}/n\mathbf{Z}$; here Lang writes Z_n for an arbitrary cyclic group of order n. Such a group is isomorphic to $\mathbf{Z}/n\mathbf{Z}$, but not in a unique way. He will apply the results of this section in §VI.8, in a situation where the cyclic group in question is the group of nth roots of unity in a field (e.g., for $n = 4$, the group of complex numbers $\{1, i, -1, -i\}$), so that it would be inconvenient to have the results stated only for the ''standard'' group.

Note that if m is an exponent of the abelian group A, then any multiple rm of m is also an exponent of A. If we work with ''standard'' cyclic groups, we have a canonical embedding $\mathbf{Z}/m\mathbf{Z} \to \mathbf{Z}/rm\mathbf{Z}$, taking $[1]$ to $[r]$, whose image is precisely the group of elements of $\mathbf{Z}/rm\mathbf{Z}$ of exponent m. If we compute $A^\wedge$ with respect to each these cyclic groups, we find that all homomorphisms of A into $\mathbf{Z}/rm\mathbf{Z}$ have range in the image of this canonical embedding, so that there is a natural way of identifying the two groups called ''$A^\wedge$'' that we get by dualizing with respect to these two cyclic groups. Unfortunately, when we work with arbitrary cyclic groups, we lose this canonical identification; thus, Lang is being a bit sloppy in writing ''$A^\wedge$'' without specifying which cyclic group is being used. A good rule for interpreting his results precisely is that whenever finitely many groups are mentioned, m denotes some *common* exponent of these groups, and duals are then taken with respect to some fixed cyclic groups Z_m of order m.

P.47, Theorem I.9.1 [>]: This theorem makes two statements of isomorphism. I claim that the first of these is a ''natural'' isomorphism, and the second an ''unnatural'' one. Indeed, the theorem gives one a precise prescription for constructing an isomorphism $(B \times C)^\wedge \cong B^\wedge \times C^\wedge$; but the isomorphism $A^\wedge \cong A$ for A a finite abelian group depends on one's choice of a family of cyclic subgroups having A as their direct sum: pick a different decomposition, and you will get a different isomorphism. We shall see later that ''natural'' constructions have certain particularly useful properties that constructions requiring choices

do not have. The isomorphisms described in the remaining two results in this section are again ''natural''.

P.48, last line [=]: ''is in fact'' should be ''is in effect'', or better, ''induces''. The same correction applies on the third line of the next page. There, the homomorphism of the second display is the map taking each coset $[x'] \in A'/B'$ (i.e., if additive notation is used, $x' + B'$, if multiplicative notation is used, $x' B'$) to the homomorphism $\psi_{x'} : A/B \to C$.

P.49, third line of Theorem I.9.2 [>]: Before the word ''*Assume*'' add the sentence ''*Then* A/B *and* A'/B' *are of exponent* m.'' This comes easily out of the proof given.

I will insert here an important result which Lang has omitted.

Proposition I.9.c1. *Let* A *be a finite abelian group, and let us define a map* $\theta_A : A \to A^{\wedge\wedge}$ *by* $\theta_A(x)(\alpha) = \alpha(x)$ $(x \in A,\ \alpha \in A^{\wedge})$. *Then* θ_A *is an isomorphism. Moreover, if* $f : A \to B$ *is any homomorphism of abelian groups, then* $\theta_B f = f^{\wedge\wedge} \theta_A$; *i.e., the diagram*

$$\begin{array}{ccc} A & \xrightarrow{\ f\ } & B \\ \downarrow{\scriptstyle \theta_A} & & \downarrow{\scriptstyle \theta_B} \\ A^{\wedge\wedge} & \xrightarrow{\ f^{\wedge\wedge}\ } & B^{\wedge\wedge} \end{array}$$

commutes.

Sketch of Proof. The fact that θ_A is a homomorphism, and the equation in the last statement (which intuitively says that the maps θ_A and θ_B identify the finite abelian groups A and B with their double duals in a way ''consistent with'' the construction $f \mapsto f^{\wedge\wedge}$ on group homomorphisms), are straightforward verifications, which you should write out; the nontrivial result is that these maps are *isomorphisms*. The most direct way to get this is to write A as a direct sum of cyclic subgroups, and verify that θ_A carries each of these summands isomorphically to the corresponding summand in $A^{\wedge\wedge}$.

An alternative way is to apply Theorem I.9.2, with $A^{\wedge}$ in the role of A, A in the role of A', Z_m in the role of C, and the evaluation function $(\alpha, x) \mapsto \alpha(x)$ in the role of the given bilinear map. Then you will find that the kernel on the left is zero (why?), hence, writing K for the kernel on the right, that Theorem says that Lang's canonical map ψ gives an isomorphism of A/K with $(A^{\wedge}/0)^{\wedge} = A^{\wedge\wedge}$. Since A and $A^{\wedge\wedge}$ have the same order, the kernel K must also be zero, so ψ gives an isomorphism $A \cong A^{\wedge\wedge}$. Moreover, comparing definitions, one sees that ψ is the map we have named θ_A. □

P.49, Proof of Corollary I.9.3 [=]: To get this as a ''special case'' of the theorem, consider the bilinear map of the last display of p.48, with $C = Z_m$, so that $\mathrm{Hom}(A, C)$ is just $A^{\wedge}$, then restrict the resulting bilinear map $A \times A^{\wedge} \to Z_m$ to a map $B \times A^{\wedge} \to Z_m$, and apply the theorem to this restriction. (The kernel of this map on the left will be trivial – why?)

P.49, end of §I.9 [>]: As we noted earlier, the dualities constructed as in this section using different cyclic groups Z_m and Z_n, where m and n are both exponents of some family of abelian groups, can be related by regarding Z_m as embedded in Z_n whenever m divides n. In fact, in some pleasant cases those embeddings are actually inclusions. This is true when Z_m is the multiplicative group of all mth roots of unity in an appropriate field, mentioned earlier. For a case accessible to us now, consider the factor group $\mathbf{Q}/\mathbf{Z}$. For each m, the subgroup of elements of exponent m is the cyclic group of order m generated by the image of $1/m$, i.e., $(1/m)\mathbf{Z}/\mathbf{Z}$, and clearly, whenever m divides n we have an inclusion of these finite groups. Thus, the various duals $A^{\wedge}$ of a finite abelian group A that one obtains using different cyclic subgroups of $\mathbf{Q}/\mathbf{Z}$ can all be written $\mathrm{Hom}(A, \mathbf{Q}/\mathbf{Z})$, yielding a single coherent duality theory.

(If you some day study topological abelian groups, you will learn of ''Pontryagin duality'' among locally compact abelian groups, defined by $A^{\wedge} = \mathrm{Hom}(A, \mathbf{R}/\mathbf{Z})$, where ''Hom'' is now defined as the

group of *continuous* homomorphisms, and is given an appropriate topology. It is easy to see that the torsion subgroup of $\mathbf{R}/\mathbf{Z}$ is $\mathbf{Q}/\mathbf{Z}$, from which it follows that for A finite, $\mathrm{Hom}(A, \mathbf{R}/\mathbf{Z}) = \mathrm{Hom}(A, \mathbf{Q}/\mathbf{Z})$. Hence Pontryagin duality, when restricted to finite groups, gives the duality treated in this section.)

We remark that Lang's symbol $A\hat{}$ is not standard. A much commoner symbol for dualities of various sorts is $\hat{A}$. (Perhaps Lang wishes to avoid this symbol because it has so many other uses, such as denoting various sorts of ''completion''.)

The interested student might like to think about which definitions and arguments in this section make sense for the constructions $\mathrm{Hom}(-, R)$ for an arbitrary abelian group R, and which are specific to cyclic groups.

Re §I.10. Inverse limit and completion.

This is an interesting subject, but one which we generally don't have time to cover in Math 250A. Some time I hope to go through this section and set down full commentary. I give below a few errata supplied by Bjorn Poonen who covered this section in Fall 2003, and some comments of my own resulting from reading the points to which those errata applied.

P.50, beginning of last line of first paragraph [=]: lim should be $\varprojlim$ (as on the preceding line); in earlier printings, the first two displays are also missing the arrow beneath the symbol.

I also suggest reading $A[p^n]$ in place of $A[p^{n+1}]$ in the first display. The limit group is the same whichever indexing is used, but the former indexing is more straightforward, and consistent with the third display.

P.51, third paragraph [>]: There are two sorts of ''limits'' of groups (or other mathematical structures), known as ''direct limits'' and ''inverse limits''. The former is constructed from what is called a ''directed system'' of groups, the latter from what is called an ''inversely directed system''. In this section, Lang only discusses the latter construction, and in this paragraph he writes ''(inversely) **directed system**'', meaning that he will often drop the word ''inversely'' for brevity. However, I would advise against doing so, since it can lead to confusion as soon as the other concept is also present; he will introduce that on p.160. (One might use an adjective to distinguish the two sorts of systems whenever there is ambiguity, and say ''directed system'' when the context indicates which sort is meant; the trouble is that I don't know of any adjective commonly used to specify the sort of system from which one constructs a direct limit of groups. Perhaps one could call it a ''forward directed system''.)

Roughly speaking, an inversely directed system is a (generally infinite) diagram of groups and homomorphisms among them, satisfying certain conditions on composition of these homomorphisms, and such that given any two groups in the diagram, there is at least one group that has arrows into both of them. One makes this precise by indexing the groups by an appropriate sort of partially ordered set I, and letting the system have a map from one group another whenever there is an inequality between the indices.

Of course, one has to decide whether the homomorphisms should go from the object with the smaller index to the one with the larger index or vice versa. Moreover, of these two possibilities, we might make the *same* choice for both kinds of limits, indexing our systems of groups by partially ordered sets with *opposite* properties, or make *opposite* choices for the two sorts of limits, indexing them by partially ordered sets with the *same* property! Different authors make different choices. Lang, at a later point where he discusses both constructions in a general setting (pp.160-161) uses in both constructions the same sort of ordered set, called a ''directed'' partially ordered set, and indexes the directed system in different ways in the two cases. The background of this choice is that the earliest cases that were considered, before the general concept was developed, involved ''sequences'' of groups, and people naturally indexed both the directed and inverse systems by the positive integers. However, be prepared to see different notations in different works. The properties of the diagrams defining ''direct'' and ''inverse'' limits will always be the

same, but the formalism of their descriptions in terms of partially ordered sets will differ.

For an additional bit of confusion, the terms ''direct limit'' and ''inverse limit'' have synonyms, ''inductive limit'' and ''projective limit'' respectively. I find these unsuggestive, and often have trouble remembering which means which. Lang generally keeps to ''direct limit'' and ''inverse limit''; I will give corrections below in the few places where he slips and uses the other terms. (The term ''profinite group'', introduced toward the bottom of p.51, is short for ''projective limit of finite groups''. That is a standard term, and when I need to remind myself which of the above names refers to which kind of limit, I work backward from that.)

P.52, last two paragraphs [<]: Lang doesn't seem to have made up his mind whether he wants to tell us about the completion of a group G with respect to an arbitrary directed system $\mathcal{F}$ of normal subgroups, or only about the special case where $\mathcal{F}$ is the set of all normal subgroups of finite index, so what he says is a bit inconsistent. The latter construction, called the ''profinite completion'' of G, is important; but since he doesn't go into the theory of that completion, let us assume $\mathcal{F}$ is a general directed system of normal subgroups. I will indicate the necessary adjustments in what he writes along with other corrections below.

P.52, first three sentences of next-to-last paragraph [=]: Lang defines a ''Cauchy family'' in G to be indexed by the directed partially ordered set of subgroups $\mathcal{F}$; but this is not in general a natural choice. A more general definition makes a Cauchy family a system $(x_j)_{j \in J}$ indexed by any directed partially ordered set J, and satisfying the condition that for every $H \in \mathcal{F}$ there exists a $j \in J$ such that for all k, $k' \geq j$ one has $x_k x_{k'}^{-1} \in H$. However, Lang has introduced the concept of Cauchy family only for perspective; he will not use the concept, so you do not have to learn one version or the other.

P.52, last two sentences of next-to-last paragraph [=]: Let us put off for the moment the question of why having $\mathcal{F}$ denumerable means that ''in practice one can work with (Cauchy) sequences'', and merely note that Lang is here saying that the system of subgroups of finite index is often countable, and is often the system we are interested in.

P.52, final display [=]: The left-hand side of the display should read $\varprojlim_i G/(H_1 \cap \ldots \cap H_i)$. Likewise, on the top of the next page, ''the sequence $\{H_i\}$'' should be ''the sequence $\{H_1 \cap \ldots \cap H_i\}$''.

The key idea here is that from a countable cofinal *set* of subgroups $H_1, H_2, H_3, \ldots$ one can obtain a countable cofinal *chain*, $H_1 \supseteq H_1 \cap H_2 \supseteq H_1 \cap H_2 \cap H_3 \supseteq \ldots$, and that the inverse limit over *that* chain is isomorphic to the inverse limit of the original system. This last statement is not obvious, but, as Lang says, it makes a nice exercise.

(Actually, one can show from the cofinality assumption that there is a *subsequence* (H_{i_m}) of (H_i) which satisfies $H_{i_{m+1}} \subseteq H_{i_m}$ for all m, and is still cofinal in $\mathcal{F}$. This gives the desired chain without taking intersections, and the statements Lang makes hold with H_{i_m} in place of H_i. This may be closer to what he had in mind, but more words would have to be inserted to express it.)

Re §I.11. Categories and functors.

P.53, definition of ''category'' [~]:

To motivate this concept it is best to begin by comparing it with some more familiar concepts.

The definition of a *monoid* is obtained by abstracting the properties of the set of all maps from a mathematical object X into itself: the set of such maps is closed under the operation of composition, which satisfies the associative law, and it contains an identity element, the composite of which with any other element f is f. A general system of elements with an operation named ''composition'', having these properties, *but not necessarily assumed to arise this way*, is called a *monoid*.

Similarly, the concept of a *group* is motivated by considering the properties of the set of *automorphisms* of a mathematical object. The set of such automorphisms has a structure of monoid, but is also closed under taking *inverses*; this gives another operation, which satisfies familiar identities.

If one considers the set of *subobjects* of a mathematical object, there are no operations one can automatically define on this set, but there is a relation of ''inclusion''. A generalization and abstraction of the properties of this relation motivates the definition of *partially ordered set*.

Next, an unfamiliar, but nevertheless similar situation: Suppose we consider two mathematical objects X_0 and X_1, and the set S of all maps among these. This set falls naturally into four pieces: S_{00}, consisting of all maps from X_0 to itself, S_{01}, consisting of maps from X_0 to X_1, etc.. One can compose various of these maps, but not every map is composable with every other, so one does not have a composition operation $S \times S \to S$, but rather, a family of 8 operations, $S_{jk} \times S_{ij} \to S_{ik}$ $(i,j,k \in \{0,1\})$. One easily sees that these 8 operations satisfy 16 associativity laws, corresponding to multiplication on the 16 sets $S_{jk} \times S_{ij} \times S_{hi}$, and that there are 2 identity elements, $\mathrm{id}_0 \in S_{00}$ and $\mathrm{id}_1 \in S_{11}$. One might abstract this situation, and call a family of four sets S_{00}, S_{01}, S_{10}, S_{11} given with operations satisfying the appropriate conditions a ''*bimonoid*''. Note that, as in the definitions of ''monoid'', ''group'' etc., this definition would *not* presuppose that the sets S_{ij} were actually sets of maps among two mathematical objects, nor that the composition operations were composition of such maps, but only that the system satisfied the algebraic laws motivated by that situation. (A virtue of this abstraction is that one could consider many ''representations'' of *the same* abstract ''bimonoid'' on different pairs of objects.)

One might go on to define ''trimonoid'', ''quadrimonoid'' etc., but this would clearly be a waste of paper: once the pattern is clear, one should make a general definition that covers all cases; namely, a model for the algebraic structure one gets when one considers a *family of mathematical objects and all the maps among these*. The resulting sort of structure is called a *category*. The family that replaces the index-set $\{0,1\}$ in the definition of ''bimonoid'', the index-set $\{0,1,2\}$ in the unstated definition of ''trimonoid'', etc., is called the class of *objects* of the category; the family that replaces the set S is called the class of *morphisms*.

One of my main reasons for giving the above discussion is to make clear that in the definition of a category, as in the preceding definitions, we do *not* assume that the structure we are considering actually consists of a family of mathematical objects of some sort and the homomorphisms among its members. The relation between an abstract category and a category arising by taking a class of mathematical objects and the homomorphisms among them is analogous to the relation between an abstract group and the group of automorphisms of a mathematical object. In each case, the abstract concept is *modeled on* the concrete concept; indeed, we might think of an abstract group as giving ''a potential structure for the automorphisms of an entity'', and of an abstract category as ''a potential structure for a class of entities and the homomorphisms among them''. But in each case, the abstract concept has less information than the motivating situation. Thus, though in talking about a group of permutations of a set one can ask a question such as, ''Does this group have any fixed points?'', in talking about an abstract group, one cannot. Similarly, in proving results about abstract categories, one cannot use arguments that begin ''Let x be an element of the object X.'' (The temptation to do so is strong for the beginner! In every class where I have introduced the concept of category, I have emphasized the above point, and yet in almost every such class, at least one student has used such an argument in his or her homework.)

P.54, end of second paragraph of **Examples** [>]: Lang says the axioms for a category are trivially satisfied by these familiar examples. The verification of the nontrivial axioms **CAT 2** and **CAT 3** is indeed trivial in these cases; but whether they satisfy ''trivial'' axiom **CAT 1** is an iffy question.

For instance, if A is the set of even integers and B the set of all integers, is the inclusion map of A in B the same as the identity map of A? By the usual definition of function that one learns in a course in set theory (as a certain sort of subset of the direct product of two sets), they are the same; hence if in $\mathbb{S}$ we define $\mathrm{Mor}(A,B)$ to be the set of all functions $A \to B$, then $\mathrm{Mor}(A,B)$ and $\mathrm{Mor}(A,A)$ are *not* disjoint, contradicting **CAT 1**.

There are various ways to fix this.

One is to slightly change our definition of function, letting a function $A \to B$ mean a 3-tuple (A, f, B), where f is what we previously called a function from A to B. Then functions from A to B in our new sense are in one-to-one correspondence with functions from A to B in the old sense, but the set of functions from A to B in the new sense is disjoint from the set of functions from A' to B', unless $A = A'$ and $B = B'$.

Another, only subtly different, is to keep our definition of function unchanged, but define a morphism from A to B in $\mathbb{S}$ to mean a 3-tuple (A, f, B) as above.

A third approach, taken in some developments of category theory, is to drop **CAT 1** from the axioms. This has some disadvantages; as one of many examples, if we are given a morphism f, the statement ''f is invertible'' is ambiguous; it only becomes precise if one says ''f is invertible as a map $X \to Y$''. Nevertheless, one may prefer to leave it out simply to minimize the gap between one's category-theoretic usage and conventional mathematical usage. (This is the decision I make in my Math 245 notes.)

We will follow the textbook and assume **CAT 1** here. I leave it to you which way you prefer to adjust your conventions to make this consistent with our non-categorical notation; it will not affect the material to follow.

(Incidentally, while in conventional usage, a set-map has a well-defined domain but not a well-defined codomain, if we leave out **CAT 1**, we have no guarantee that either the domain or the codomain of a morphism in a category will be well-defined.)

P.54, remaining Examples (after $\mathbb{S}$, **Grp**, **Mon**) [<]: I recommend skipping the material from here through p.57, except for the two paragraphs near the top of p.55 beginning ''Next we consider ...'' and ''More generally ...''. (In the fourth line of the latter paragraph, for ''In practice'' read ''In many cases''.) Most of Lang's other examples involve material that few beginning students will have had, and are not clearly presented. (If you do continue on p.54, you will find that his comment three lines later, ''It is important to emphasize ...'', is quite off the wall, since none of the three examples he has given so far had the property that its morphism-sets formed groups.)

In place of (or in addition to) Lang's examples, read the examples and (optionally) the notes on category theory and set theory below, then go back to Lang where indicated at the end of those notes.

Examples. Here are a couple of interesting categories in which the morphisms are not set-maps. First, one can define a category **Rel** whose objects are the same as the objects of **Set** (what Lang calls $\mathbb{S}$; most authors use more mnemonic names), but where a morphism $X \to Y$ means any *relation* $R \subseteq X \times Y$, and where composition is defined to mean composition as relations; i.e., for $R \subseteq X \times Y$ and $S \subseteq Y \times Z$, one defines

$$SR = \{(x, z) \in X \times Z \mid (\exists y \in Y)\ (x, y) \in R, (y, z) \in S\}.$$

Secondly, for students who have seen the definition of the fundamental group $\pi_1(X)$ of a topological space X, observe that given such a space X, we may define a category $\boldsymbol{\pi}_1(X)$ in which the objects are the points of X, a morphism from the point x_1 to the point x_2 means a homotopy class of paths from x_1 to x_2, and homotopy classes of paths are ''composed'' as in the definition of the fundamental group. (A curious property this category is that every morphism is invertible. A category with this property is often called a *groupoid*.)

Another class of categories is described in exercise I.11:6(a).

Category theory and set theory (optional). If one wants to study concepts like ''the category of all groups'', one is faced with the problem that under conventional set theory, all groups form a *class*, but not a *set*, and there is very little one is allowed to do with classes. One often calls a category which forms a genuine set a ''small category'', and though one can work easily with these, they exclude most of the categories one is interested in.

Here is a creative way out of this dilemma, due to Alexander Grothendieck. Conventional set-theory

(ZFC, i.e., Zermelo-Fraenkel set theory with the axiom of Choice) has various axioms asserting that given certain sets, one can obtain certain other sets. Let us think of these as saying that the *class of all sets* considered in the set theory is closed under certain constructions. ZFC neither asserts nor denies that there may be *sets* $\mathbb{U}$ within the theory that are themselves closed under all these constructions (i.e., such that, for all $X \in \mathbb{U}$ every member of X belongs to $\mathbb{U}$, the power set of X belongs to $\mathbb{U}$, etc., etc.). Let us call such a set $\mathbb{U}$ a ''universe'' (or a ''Grothendieck universe''). Thus, a universe will look *internally* like a full ZFC set theory, though externally it will be a *member* of our set-theory. Let us now adjoin to the axioms of ZFC one further axiom, the ''Axiom of Universes'', saying that every set is a *member* of a universe. Given a choice of a particular universe $\mathbb{U}$, let us define a ''small set'' (relative to $\mathbb{U}$) to mean a member of $\mathbb{U}$, and then define a ''small group'' (or other mathematical object) to mean a group (etc.) which is small as a set. Within our larger set theory, there exists a set of *all small* sets (namely $\mathbb{U}$), hence also a set of all small groups, and similarly for other sorts of mathematical objects. Thus, in this set theory, the categories of all small sets, small groups, etc., are genuine sets (though not small sets!), and can be handled mathematically without kid gloves. (Even the collection of all categories of the above sort – categories having object-sets and morphism-sets which are subsets of $\mathbb{U}$ – will form a genuine set in our set-theory. It will not be a small set with respect to $\mathbb{U}$, but it will be small with respect to any larger universe $\mathbb{U}'$.)

Since ZFC was apparently sufficient for studying group theory before categories came along, one can regard the category of *small* groups (with respect to an arbitrary fixed universe $\mathbb{U}$) as embracing the subject of traditional group theory, and likewise for other sorts of mathematical objects; and these categories can be treated conveniently within set theory. If you like this approach, you should henceforth read references to ''the category of all groups'', etc., as referring to these categories.

We will not develop these ideas further in this course, but simply note that they give a way out of the set-theoretic dilemma with which we began this remark. I may, in fact, for consistency with common usage, sometimes speak of the objects of the category of groups as forming ''the class of all groups''. Lang, as you may have noticed, begs these questions by using the vague term ''collection'' in his definition of category on p.53. (In my 245 notes, I start with a similarly vague ''Provisional Definition'' of category when I motivate the concept in §6.1, but then introduce the Axiom of Universes in §6.4, revise the definition of a category to require that its objects and morphisms form sets, and thereafter use this modified set theory.)

Now return to Lang, at the paragraph on p.57 beginning ''Let $\mathscr{A}$ be a category ... ''.

P.57, line after square diagram, ''... Strictly speaking ...'' [>]: Recall the comment on **CAT 1** in my discussion of p.54. If $\mathscr{A}$ satisfies **CAT 1**, then defining morphisms of $\mathscr{C}$ as tuples (f, φ, ψ, f') would insure that $\mathscr{C}$ satisfied that condition. Lang's approach to the question seems to be that in theory, we make that definition, but when there is no danger of ambiguity, for the sake of brevity we just write (φ, ψ).

P.58, second paragraph of Examples [~]:

Some notational matters.

Lang wants to show us here that a ''free abelian group'' is an example of a universal repelling object. To understand this approach, recall that the universal property of the free abelian group $F_{\mathrm{ab}}(S)$ on a set S says that it is an abelian group with a map f of S into it, which is ''universal'' among all abelian groups with maps of S into them, in the sense that such that given any abelian group B and map $g\colon S \to B$, there is a unique group homomorphism $g_*\colon F_{\mathrm{ab}}(S) \to B$ respecting maps from S; i.e., such that $g(s) = g_*(f(s))$ for all $s \in S$. (Lang, p.38, last paragraph, ''If g: ...''.) This suggests that, given a set S, we should define a *category* whose objects are abelian groups given with maps of S into them, with morphisms defined as group homomorphisms which respect these maps (i.e., which form commuting triangles with them). The group $F_{\mathrm{ab}}(S)$, given with the map f, will then be a ''universal repelling

object'' in this category.

The straightforward way to formalize the definition of this category takes for the objects the pairs (B, g) such that B is an abelian group and $g\colon S \to B$ a set-map. However, note that if such a map g is considered to intrinsically specify its codomain B, then g alone gives as much information as the ordered pair (B, g). Hence Lang, and many other authors, parsimoniously define the *objects* of their auxiliary category simply to be these maps g. Lang then forgets to mention that he is redefining the ''free group on S'' to mean the object of this category which we would call $(F_{\mathrm{ab}}(S), f)$, and which he simply writes f. Once this redefinition is noted, what he says is correct.

In fact, he will not be consistent – he will sometimes speak of the free abelian group as a map (as above), sometimes as a group given with such a map, and sometimes simply as a group. My own advice is to avoid the ''just a map'' definition, and speak of free abelian groups either (precisely) as certain ordered pairs (A, f), or (loosely) as certain groups A.

Part of the difficulty is more than just terminological. Though the basic concept of a ''free abelian group'' is that of an abelian group given with a map of a specified set into it and having an appropriate universal property (equivalently, an abelian group in which a specified family of elements forms a basis), we are sometimes interested instead in a group which merely has the property that there *exists* a map of some (unspecified) set into it with that property; i.e., which has *some* basis. For instance, when we say that ''a subgroup of a free abelian groups is free'', we are in the later context. So we not only have three closely related *formalizations* to choose between (object, object-with-morphism, and morphism alone), but also two closely related *concepts* (abelian groups free on specified bases, and abelian groups free on unspecified bases). However, if you keep the possible formalizations and concepts in mind, you should generally be able to see which of these an author is using, and hence what he or she means. Whenever the *universal property* of free abelian groups is being discussed, the concept in question is the one involving a specified basis.

In the immediately following passage in Lang, on products and coproducts, you will encounter a similar situation. The *product* P of two objects A and B (to take the case you have some familiarity with) will be defined to be an object given with morphisms f, g to A and B respectively, called ''projections'', such that P is universal among all objects C given with two morphisms $\varphi\colon C \to A$ and $\psi\colon C \to B$. Lang will characterize the product as the universal attracting object (P, f, g) in the category of all such 3-tuples (C, φ, ψ). (Following the same principle as for free abelian groups, he could take the objects of this category simply to be the pairs (φ, ψ). Mercifully, he does not.) But he subsequently vacillates, sometimes using the term ''product'' for the 3-tuple (P, f, g) and other times for the object P.

These difficulties are not a specialty of category theory. As we noted in our comments on §I.1, when we speak of ''a monoid G'' we should, more accurately, specify its operation, and write ''the monoid (G, μ)''. But since category-theory is quite new to most students at this level, these variations in preciseness are particularly noticeable.

As a much more petty distinction, note that in setting up a notation for the entity given by the product object and its projection maps, one has a choice between using 3-tuples (P, f, g) and ordered pairs $(P, (f, g))$. Why might one use the latter? To make it equally convenient to define the product P of an arbitrary family of objects $(A_i)_{i \in I}$, for which the natural notation is $(P, (f_i)_{i \in I})$. There is no absolute answer as to which notation should be used; but you should be aware that Lang uses both forms.

Incidentally, what Lang calls ''universal repelling'' and ''universal attracting'' objects in categories are usually called ''initial'' objects and ''terminal'' (or ''final'') objects respectively.

P.59, first **Example**, second line [=]: By the ''cartesian product'', Lang means the ordinary set-theoretic product, i.e., the set of all I-tuples (a_i) with $a_i \in A_i$. (It is named after René Des**cartes**, who introduced the idea that points of the plane could be described by ordered pairs of real numbers.)

P.59, next paragraph [~]: Lang says that the symbols $A \times B$ and $\prod A_i$ are used to denote products in a general category $\mathscr{C}$. This is correct if we understand ''product'' to mean the universal *object* of $\mathscr{C}$; if

we mean the universal *tuple*, then these symbols denote its first component, and the whole tuple will have the form $(A\times B, f, g)$ or $(\prod A_i, (f_i))$. (I have here copied the symbols Lang is using for the projection maps; these vary from author to author; p_A, p_B, p_i $(i\in I)$ etc., or π_A etc. are frequently used.)

The same comments apply to the symbol $\coprod$ for ''coproduct'', introduced on the next page. (Convenient symbols for the coprojection maps associated with a coproduct are q_A, q_B, q_i etc..)

P.60, last **Example** [<]: Ignore this unless you have seen tensor products of rings. (These are developed in §XVI.6 of Lang; though they could easily be introduced in Chapter II.)

P.61, three lines following the triangular diagram [=]: Replace these lines by: ''The **product** in $\mathcal{C}_Z$ of two objects $f: X \to Z$ and $g: Y \to Z$ is called the **fibered product** of X and Y (or of f and g) over Z, and is denoted $X\times_Z Y$. More precisely, this term describes the 3-tuple $(X\times_Z Y, p_1, p_2)$, where p_1 and p_2 are the canonical maps to X and Y. This 3-tuple is universal for making a commuting diagram:''

Incidentally, why the term ''fibered''? Given a set X *over* (i.e., with a map into) a set Z, the inverse image in X of each element $z\in Z$ is called a ''fiber'' of the map $X \to Z$. (To see why, draw a picture in which Z is shown as a line, and X as a ''glob'' above it, projected ''down'' to it by the given map; and sketch the inverse images of a few points.) The ''fibered product'' P of two sets X and Y over Z is so called because for each $z\in Z$, the fiber of P at z is the direct product of the fibers of X and Y at z.

The term ''fibered'' is sometimes extended, as Lang will do on the next page, to the dual case of coproducts, though there is no intrinsic motivation for it there. (A name more often used in that case is ''the coproduct of X and Y with amalgamation of Z''.)

P.61, after the first diamond-shaped diagram [<]: The next 8 lines (starting with the italicized line ''*Fibered products and coproducts* ...'') should go on next page, just before the word **Functors**. One should also say what map to Z is used to make this subgroup an object of $\mathcal{C}_Z$. It is the map taking each pair (x, y) satisfying the indicated equation to the common value in Z of the two sides of that equation.

P.62, **Example** just before the heading **Functors** [=]: After ''their product'' add ''in $\mathcal{S}^Z$''.

P.62, definition of functor [>]: If we think of a category as a type of algebraic object, then *functor* is the name for a *homomorphism* among such objects. On the other hand, if we think of the typical category as ''a class of mathematical objects, and the morphisms among them'', then we find that functors generally correspond to *constructions* that build one sort of mathematical object from another. Indeed, in most such mathematical constructions, it turns out not only that for every object X of one sort one gets an object $F(X)$ of the other sort, but also that for every morphism $a: X \to Y$ of objects of the first sort one gets a morphism of constructed objects, $F(a): F(X) \to F(Y)$ (or sometimes $F(Y) \to F(X)$). Although these constructed *morphisms* may not be noticed at first, they eventually turn out to play as important a role in the theory as the constructed *objects*.

For instance, you have seen the construction of the free abelian group on a set. You should now verify that given any set-map $a: X \to Y$, there is a natural way to construct a group homomorphism $F_{\mathrm{ab}}(a): F_{\mathrm{ab}}(X) \to F_{\mathrm{ab}}(Y)$. Similarly, given a homomorphism $G \to H$ of abelian monoids, there is a natural way to get a map $K(G) \to K(H)$ of their Grothendieck groups. In each case, this can be verified from the construction, or as an easy application of the universal property. In each case, also, it is not hard to show that these constructions respect identity maps and composition of maps.

The class of functors which Lang calls ''stripping functors'' are called by essentially all other authors ''forgetful functors'', and we shall use the latter term here.

There is a trivial but important class of functors which Lang does not mention. Any category $\mathcal{C}$ has an *identity functor* $\mathrm{Id}_{\mathcal{C}}: \mathcal{C} \to \mathcal{C}$, which you should have no trouble defining precisely.

Incidentally, Lang defines a functor as a ''rule'' rather than a function for the same reason he defines a category as a ''collection'' rather than a set. Cf. note on *Category theory and set theory* above.

P.63, starting with second example [<]: The remaining examples on this page refer mostly to nonalgebraic material, and so should be skipped by students not familiar with it. (The second example

refers to a construction defined in §I.10, and so should likewise be skipped if that section was not covered.) I suggest resuming the reading of Lang with the second example on p.64.

P.64, middle [>]: The standard symbols for the functors that Lang denotes M_A and M^B are h_A and h^B, and we shall use these here. The objects A and B respectively are called the *representing objects* for these functors.

P.65, middle [>]: I advise skipping the remainder of this page. (But if you read it, note that in addition to Grothendieck's work, a very important contribution to the study of what Lang calls ''representation functions'' was Freyd's paper *Algebra-valued functors in general and tensor products in particular,* Colloquium Mathematicum (Wrocław) **14** (1966) 89-106.)

Below are further remarks on the material of §I.11.

Further notes on functors.

I have suggested two ways of thinking about functors: as ''homomorphisms'' between categories, and as abstractions of what generally happens when we have a way of constructing one sort of object from another. Here is a third important viewpoint (or perhaps a refinement of the first): A functor from a category $\mathscr{A}$ to a category $\mathscr{B}$ can be thought of as a system of objects of $\mathscr{B}$, parametrized by the set of objects of $\mathscr{A}$, and tied together by a system of morphisms, likewise parametrized by the morphisms of $\mathscr{A}$, which satisfy composition relations corresponding to those satisfied in $\mathscr{A}$. A *morphism of functors* can then be thought of as a way of setting up morphisms from the *objects* of one of these ''systems'' to the corresponding objects of the other, which respects the connecting morphisms. Note that the existence of a concept of morphism of functors $\mathscr{A} \to \mathscr{B}$ makes the set of functors $\mathscr{A} \to \mathscr{B}$ itself into a category, often denoted $\mathscr{B}^{\mathscr{A}}$. An easy illustration is given in Exercise I.11:3. We will note more examples in our comments on the next section.

Lang has only indicated one class of examples of *contravariant* functors so far, the functors he calls M^B (p.64), and which we will write more standardly as h^B, defined by $h^B(A) = \mathrm{Mor}(A, B)$; but this is an extremely important class. You should draw diagrams to convince yourself that given a fixed object B and a morphism $X \to Y$, one does indeed get a morphism $h^B(Y) \to h^B(X)$, and that this construction respects composition and identity morphisms. Some examples: If $\mathscr{A}$ is the category of vector spaces over a field k, and B is the 1-dimensional space k, then h^B is the ''dual vector space'' construction. The behavior of this construction with respect to linear maps probably constitutes the first example of contravariance most students see, though, of course, they do not see it under that name. Here the hom-sets can be given structures not only of sets, but of vector spaces, giving us a functor from vector spaces to vector spaces. (The circumstances under which the sets arising as values of a representable functor, covariant or contravariant, acquire algebraic structure, are something that we won't be able to cover here. It is treated in §§9.1-9.3 of the 245 notes. However, we will note in passing some further cases in the next few examples below.) If $\mathscr{A}$ is the category of abelian groups of exponent n, then the contravariant functor $h^{\mathbb{Z}_n}$, from this category to **Set**, can likewise be made a functor to abelian groups of exponent n; this is the duality developed by Lang in §I.9. If $\mathscr{A} =$ **Set** and B is the object $2 = \{0, 1\}$, note that each object $h^B(X)$ can be identified with the set of all subsets of X, by the correspondence between subsets of X and their characteristic functions; this is the source of the notation 2^X for the set of all subsets of X. If $f\colon X \to Y$ is any map, then the induced map $2^Y \to 2^X$ is the ''inverse image under f'' construction. (The ''set of all subsets'' construction can also be made a *covariant* functor, by using images rather than inverse images. Both functors are important, but the contravariant functor has better properties, and is the more important.) We remark that this construction be made a functor to *Boolean rings*.

For an example of a *covariant* representable functor, let us fix any integer n, and consider the construction associating to every group G the set of its elements of exponent n, which we might write $E_n(G) = \{x \in G \mid x^n = e\}$. It is easy to see that a group homomorphism $G_1 \to G_2$ will carry elements

of exponent n in G_1 to elements of exponent n in G_2, making E_n a functor **Group** $\to$ **Set**. I claim that this functor is in fact *isomorphic* to h_{Z_n}. Intuitively, the idea is that to determine a homomorphism from Z_n to a group G, one simply has to specify where the generator $[1] \in Z_n$ is to be sent, and the possible choices are precisely the elements of exponent n, so the set of homomorphisms $Z_n \to G$ is in natural bijective correspondence with the set of such elements. To formalize this, we define a *morphism of functors* $h_{Z_n} \to E_n$ taking each $f \in h_{Z_n}(G)$ to $f([1]) \in E_n(G)$, a morphism the other way taking each $x \in E_n(G)$ to the group homomorphism $[r] \mapsto x^r$, and observe that these are inverse to one another, hence constitute an *isomorphism* of functors. This says that the functor E_n: **Group** $\to$ **Set** is *representable* (Lang, p.65), with representing object Z_n.

As another example, if X is a set, and we define a functor S_X on the category **Ab** of abelian groups taking each group A to the set of all X-tuples of elements of A, then we find that S_X is representable, with representing object the free abelian group on X. Indeed, this is equivalent to the universal property of that group! Finally, suppose L is an abelian monoid, and we let N_L denote the construction associating to each abelian group A the set of monoid homomorphisms $L \to A$. It might seem that we should call this functor **Ab** $\to$ **Set** ''h_L'', but this is not quite correct, because L is not an object of **Ab**. However, there *is* an object of **Ab** which represents this functor – the Grothendieck group $K(L)$, again, by inspection from its universal property.

The above observations are obviously somehow related to the characterizations of Z_n, of free abelian groups, and of Grothendieck groups in terms of initial (''universal repelling'') objects: the group Z_n, with a cyclic generator specified, constitutes the initial object in the category of groups given with a distinguished element of exponent n, the free abelian group on X is initial in the category of abelian groups with an X-tuple of distinguished elements, and $K(L)$ is initial in the category of abelian groups given with monoid homomorphisms of L into them. In the Math 245 notes, the relationship between these viewpoints is developed in §§7.1-7.2. Here, let it suffice that you see that each of these examples illustrates both concepts.

Remark on the category of categories (optional). Though an ordinary category involves two sorts of entities, ''objects'' and ''morphisms'', when one studies the category of *categories*, one has three levels: categories, functors, and morphisms of functors. One can write down various laws relating composition of functors with composition of morphisms between functors, and abstract the resulting structure on the category of categories. Systems with such structure are studied under the name *2-categories*, but we will not look at them in this course. Cf. Mac Lane's *Categories for the Working Mathematician* (or for a brief discussion, §§6.10-6.11 of the 245 notes).

Notes on varying conventions.

An author writing about categories has to make certain choices of convention. Lang has chosen to write morphism-sets as $\mathrm{Mor}(X, Y)$; a more common notation is $\mathrm{Hom}(X, Y)$ (even though one says ''morphism'' rather than ''homomorphism''). When one wants to specify the category $\mathcal{C}$ unambiguously, one writes this morphism-set $\mathrm{Hom}_{\mathcal{C}}(X, Y)$; a simpler notation, also common, is $\mathcal{C}(X, Y)$. The roles of X and Y in these symbols are also occasionally reversed, i.e., some authors use one or another of these symbols to denote the set of morphisms from Y to X rather than the set of morphisms from X to Y! Independent of this choice is the choice of whether to write gf (as we do in this course) or fg for the composite of a morphism $f: X \to Y$ and a morphism $g: Y \to Z$.

We have also noted that developments differ in whether they assume the axiom **CAT 1** (morphism-sets are disjoint).

In the above matters I shall follow the choices Lang has made. On the other hand, as mentioned earlier, I will not follow Lang where he has introduced various pieces of very nonstandard terminology and notation; thus, what he calls ''universal repelling'' and ''universal attracting'' objects of a category we shall call ''initial'' and ''terminal'' objects, what he calls ''stripping'' functors we shall call ''forgetful

functors'', and the covariant and contravariant functions that he writes M_A and M^A we will denote h_A and h^A. I have not seen his term ''representation functors'' used for these; they are generally called the covariant and contravariant *hom-functors* determined by the object A. (However, the term ''*representable* functor'' for a functor *isomorphic* to one of these is standard.)

Re §I.12. Free groups.

In this section, Lang introduces the construction of the free group, using the language of category theory. My own preference is to first present this construction as a solution to a problem posed in purely group-theoretic terms, then to develop a number of similar constructions, and finally to introduce the concept of category, showing how that allows us to unify what we have been doing. Since Lang is our text, we will follow his order. Students interested in understanding better the concept and construction of free group will find a many-sided motivated development in Chapter 2 of my Math 245 notes (for which the prerequisite is the short Chapter 1). Chapter 3 of those notes treats a series of other universal constructions, which can be discussed more briefly with the ideas from Chapter 2 as background. In particular, the first eleven sections of that Chapter concern constructions involving only groups and (occasionally) monoids.

P.67, third from last line [=]: Before ''We let F'' add ''Now given g, F_0 as in the preceding paragraph''.

P.68, middle [>]: Lang has constructed the free group, but given us no idea what it looks like! We can ''understand'' it, in the sense that we can see from the universal property that in the free group on n generators $x_1, \dots, x_n$, the only equations satisfied by these n elements are the equations that are satisfied by *any* n elements in *any* group. (Note that the analogous statement is true of the free *abelian* group on n generators, with ''any abelian group'' in place of ''any group''.) We will, however, be able to deduce a very explicit description of the free group from results Lang will prove on coproducts of groups later in this section.

P.68, last paragraph [<]:

Presentations of groups by generators and relations.

Suppose G is a group, S a generating family for G, and T a set of *relations* satisfied by the elements of S in G – that is, equations $u = v$, where u and v are group-theoretic expressions in an S-tuple of variables, which are satisfied by the given family of generators of G. Then we say that G is the group *presented by the generators* S *and relations* T if the relations constituting T imply *all* relations that hold among the elements of the generating set S. This is equivalent to saying that G is the initial object (''universal repelling object'') in the category of groups given with S-tuples of elements satisfying the system of relations T. For example, our characterization of the group Z_n in our ''*Further notes on functors*'' above is equivalent to saying that it can be presented by one generator x and one relation, $x^n = e$.

In fact, for every set S, and family of equations T in an S-tuple of group-theoretic variables, there exists a group with the above property. We may see this as follows. Form the free group $F(S)$. For each relation ''$u = v$'' in T, evaluate the expression uv^{-1} in $F(S)$ using the universal S-tuple of elements of $F(S)$, and let $R \subseteq F(S)$ denote the set of elements (''relators'') obtained in this way from the relations comprising T. Let $N \lhd F(S)$ be the normal subgroup generated by R. Then $F(S)/N$, with the map of S into it induced by the canonical map $S \to F(S)$, will be the desired object. You should find its universal property straightforward to prove using the universal property of the free group, and the universal property of a factor group. This group, which as noted above is called the group ''presented by generators S and relations T'', is often written $\langle S \mid T \rangle$ or $\langle S \mid R \rangle$ by group-theorists. (E.g., the group presented by generators x, y and the one relation $x^2 = y^3$ is sometimes written $\langle x, y \mid x^2 = y^3 \rangle$ and sometimes $\langle x, y \mid x^2 y^{-3} \rangle$.)

Note that to show that a given group G has presentation $\langle S \mid T \rangle$, one must verify not only that the

generators S satisfy the relations T in G, but that the relations T imply *all* the relations satisfied by S in G. We do not have time in this course to examine how this is done in various situations (it can depend on the sense in which G is ''given''), nor the reverse problem of obtaining concrete descriptions of groups presented by generators and relations. Exercise I.12:2 gives a little practice; you might also turn back to p.9 in Lang, and see whether you can show that the two sets of equations displayed near the bottom of the page are indeed the relations in presentations of the two nonabelian groups of order 8. (In the second system of equations, note that k and m are abbreviations for ij and i^2, as indicated on the preceding line.) The student interested in learning more can turn to any of the various books whose titles are, approximately or exactly, *Combinatorial Group Theory*, the best known being the book of that title by Magnus, Karrass and Solitar. There is also a lively book, *Presentations of Groups*, by D. L. Johnson (London Mathematical Society Student Texts, vol.15, 1990) specifically on this topic, aimed at advanced undergraduates and beginning graduate students.

It is important to note the difference between relations and *identities*. E.g., the group with presentation $\langle a, b \mid a^3 = e,\ b^3 = e \rangle$ does not satisfy the identity $x^3 = e$: in order for a group generated by elements a and b to satisfy this identity, not only must the cubes of those generators equal e, but also the cubes of all other elements, e.g., $(ab)^3$, $(a^2b^2a^{-1})^3$, etc., and the two relations $a^3 = e$ and $b^3 = e$ do not imply these others, as may be seen by considering the elements $a = (123)$ and $b = (234)$ of A_4. (On the other hand, the nonabelian group of order 27 not containing a cyclic subgroup of order 9 found in exercise I.6:2 does satisfy this identity.)

Lang's terminology and notation differ in some small ways from what I have described: He writes ''determined by generators and relations'' in place of the standard ''presented by generators and relations''. He also starts off (on p.68) only considering relations of the form $r = e$, and calls the elements r the ''relations''. (The standard term is ''relators''.) Then, on the next page, he passes to relations written as equations, as discussed above.

P.69, **Example** in the middle of the page [<]: I recommend skipping everything from this example through the short paragraph **Further Examples** on p.70. Resume reading with Proposition I.12.3 on that page.

If you do read the examples mentioned, note that the ''exercise'' referred to on the line following the first display is not trivial; cf. Exercises I.12:7 and I.12:8 in this Companion. And on the next page, two lines above the first display, ''*the free group with* ...'' should be ''*the group presented by* ...''. (If we think of a free group on a set X as a group generated by an X-tuple of elements that satisfy no relations other than those implied by the group identities, and a group presented by generators and relations X and R as a group generated by an X-tuple of elements that satisfy no relations other than those implied by the relations in R and the group identities, we can understand why the latter was sometimes called ''the free group with those generators and relations''. But ''free'' now has the specific meaning defined at the bottom of p.66, and should only be used in that sense.)

P.71, Example, ending just before Proposition I.12.4 on the next page [<]: I suggest skipping this. (It is a continuation of one of the examples on p.69 that I suggested you skip.)

P.73, proof of Proposition I.12.5 (continued on next page) [~]: Lang gives the ''natural'' but messy proof of this result. There is a much neater one, due to van der Waerden, based on remembering that the group concept arises from looking at the properties of *permutations of sets*. Given two groups A and B, let us follow Lang in forming the set of sequences $(a_1, \ldots, a_n)$ (including the empty sequence), such that

(c14) All a_i lie in $(A - \{1\}) \cup (B - \{1\})$, and no two successive terms a_i, a_{i+1} lie in the *same* group A or B.

But rather than calling this set $A \circ B$ and putting a group structure on it, let us call it X, and construct a group of *permutations* of X.

This group will be generated by the homomorphic images of A and B corresponding to an action of each of these groups on X. The goal is that $(a_1, \dots, a_n) \in X$ should be the sequence arising by starting with the empty sequence $() \in X$, applying a_n to this, then a_{n-1} to the result, etc.. This tells us how we ought to define certain cases of the actions of A and B on X, namely

(i) If $(a_1, \dots, a_n)$ is an element of X, and $a \in A \cup B$ is a nonidentity element, and does not come from the same group A or B from which a_1 comes (which we take to be satisfied trivially if $n=0$), we define $a(a_1, \dots, a_n)$ to be $(a, a_1, \dots, a_n)$, which we see is indeed an element of X.

There are three other possible relations between elements $(a_1, \dots, a_n) \in X$ and $a \in A \cup B$:

(ii) If $a = 1$, we of course define $a(a_1, \dots, a_n) = (a_1, \dots, a_n)$.

(iii) If $a \neq 1$ and this belongs to the same group as a_1, we define $a(a_1, \dots, a_n)$ to be $(a\,a_1, \dots, a_n)$ if this is permitted, i.e., if $a\,a_1 \neq 1$.

Finally

(iv) If $a \neq 1$ and belongs to the same group as a_1, and if $a\,a_1 = 1$, then we let $a(a_1, \dots, a_n) = (a_2, \dots, a_n)$ (understood to mean $()$ if $n=1$).

Restricting attention to the action of elements $a \in A$, we find that the map $A \times X \to X$ given by the above ''multiplication'' is indeed a group action. (There are several cases to consider in verifying that $a(a'x) = (aa')x$, but each is a quick computation.) Similarly, the map $B \times X \to X$ is an action of B. We define $A \circ B$ to be the group of permutations of X generated by the images of these two actions.

Now if we write $\bar{a}$ for the image of $a \in A \cup B$ in Perm(X), then, simply from the fact that the group $A \circ B$ is generated by the images of A and B under certain homomorphisms $a \mapsto \bar{a}$, it is not hard to verify that every element of $A \circ B$ can be reduced to an expression $\bar{a}_1 \dots \bar{a}_n$ satisfying (c14). (Lang speaks of elements $\neq 1$ as having this form; I also consider ''1'' the case of this expression where $n=0$.) Moreover, given such an element $\bar{a}_1 \dots \bar{a}_n \in A \circ B$, it is easy to verify by induction that on applying this element to the empty sequence $() \in X$, we get the sequence $(a_1, \dots, a_n) \in X$. (The assumption (c14) guarantees that at each step of the induction we will be in case (i) above.) Hence a product $\bar{a}_1 \dots \bar{a}_n \in \mathrm{Perm}(X)$ satisfying (c14) uniquely determines its sequence of factors, so permutations with distinct expressions of this form must be distinct. We may finish the proof, as Lang does, by observing that ''identifying A and B with their images in $A \circ B$, we obtain a proof of the proposition'' (though we should discuss in class how to make that argument precise).

Although the above proof involves less messy computation than the one in Lang, there is still a step that is messier than it should be: verifying that the maps $A \times X \to X$ and $B \times X \to X$ described are really group actions, since we have to consider several cases, depending on whether our element of X begins with an element of A or an element of B, and whether the various products we get are 1 or not. But there is even a neat way around this: it involves setting up a little more machinery, but saves us work in the end. The trick is to form two bijective of copies of X, which we shall call X_A and X_B, one convenient for describing the action of A and the other for describing the action of B. Namely, given $x \in X$, define x_A to be the same element x if this begins with a member of A, while if $x = (a_1, \dots, a_n)$ where either $a_1 \in B$ or $n=0$, we define $x_A = (1, a_1, \dots, a_n)$. Thus every element of X_A begins with an element of A, and we can define the action of A on this set by in *all* cases letting $a(a_1, \dots, a_n) = (a\,a_1, \dots, a_n)$. The desired condition $a(a'x_A) = (aa')x_A$ is now immediate from the associativity of A! We define X_B and the action of B on this set in the exactly analogous manner. Because we have bijections $X_A \longleftrightarrow X \longleftrightarrow X_B$, these actions induce actions of A and B on the set X. The descriptions of these actions on X are easily seen to be those that we began with; the advantage of this approach is that the tedious verification that they *are* actions has been circumvented. This gives the most elegant proof of the proposition.

We have described the construction of a group $A \circ B$ from *two* groups, A and B, only for convenience. There is a minor way in which the two-group case is special: in any sequence of elements satisfying (c14), elements of $A - \{1\}$ and of $B - \{1\}$ occur alternately. If we had stated (c14) in terms

of this alternation, we would have to modify this statement to get the general result; but since we did not, we can say that exactly the same argument shows that given any family $(A_i)_{i \in I}$ of groups whose pairwise intersections are $\{1\}$, there exists a group $\circ_I A_i$ containing all the A_i, such that all products $a_1 \dots a_n$ satisfying (c14) are distinct.

P.74, Corollary I.12.6 [~]: Note that merely because the group $F(S)$ is generated by the elements of S, every element of this group can be written in the form $x_{i_1}^{\nu_1} \dots x_{i_r}^{\nu_r}$ as described, for some $x_1, \dots, x_n \in S$ and some family of exponents. This corollary tells us that every element of $F(S)$ has a *unique* expression in this form. This is the explicit description of the free group promised earlier; it allows us to compute easily in such groups.

P.74, Corollary I.12.8 [~]: Here Lang finally gets to the point of the construction $G_1 \circ \dots \circ G_n$. In fact, that notation of his was purely temporary; the group in question is written $G_1 * \dots * G_n$ by most group theorists and called their ''free product'', while workers with a more category-theoretic orientation call it the coproduct of groups.

The notation $\amalg X_i$ for the coproduct of an arbitrary family of objects of a category is standard, but there is no standard notation for the coproduct of two or more ''listed'' objects. I agree with Lang that $X_1 \amalg \dots \amalg X_n$ is a good choice.

In stating this corollary, Lang refers to arbitrary groups $G_1, \dots, G_n$ forgetting to mention the condition of intersecting only in 1 that he had in Proposition I.12.5. Actually, this condition fails to hold in many situations of interest, so let us note how one gets around it. Given *any* family of groups (which may even include repetitions of the same group), observe that one can achieve this disjointness condition in appropriate *isomorphic copies* G_i' of these groups. Then the same argument given in Lang shows that the map $G_1 \amalg \dots \amalg G_n \to G_1' \circ \dots \circ G_n'$ is an isomorphism; equivalently, that $G_1' \circ \dots \circ G_n'$ with the appropriate maps of the G_i's into it is a coproduct of the G_i's. (And the same is true for coproducts $\amalg_I A_i$ of not necessarily finite families of groups.)

P.74, bottom [>]: Here are some further notes on topics in the above section.

General heuristics for constructing universal objects.

A category for which one wants to find an *initial* object will typically consist of mathematical objects of some standard sort, together with certain maps of fixed objects into them having particular properties. (Examples: Groups with maps of a fixed set X into them; abelian groups with homomorphisms of a fixed nonabelian group G into them; abelian groups with monoid homomorphisms of a fixed abelian monoid M into them; arbitrary groups with homomorphisms of two fixed groups A and B into them.) Similarly, a category for which one wants to find a terminal object typically consists of objects of a certain kind, together with maps of those objects *into* certain fixed objects, having specified properties. (Examples: Groups or sets with homomorphisms into two specified groups or sets A and B; groups with monoid homomorphisms into a fixed monoid M.) In each case, it is useful to begin the search for a universal object by thinking about a general (rather than a universal) object U with the indicated extra maps, and asking, ''What handle can we get on the structure of U in terms of these maps?'' In the first type of case, this ''handle'' typically consists of specifications of certain elements of U (namely, images of elements under the given maps, and other elements obtained from these by the operations), and relations among these. If one constructs abstractly an object consisting of just such a family of elements, subject only to the relations that must hold in every case, one gets, in favorable cases, a *universal* object with the indicated type of structure, corresponding to an initial object of the category in question. (For instance, the expressions one can form from an S-tuple of elements of a general group are precisely all expressions $x_{i_1}^{\nu_1} \dots x_{i_r}^{\nu_r}$ in finite families $x_1, \dots, x_n$ of such elements, and each such expression can be *reduced* to one in which successive elements x_i, x_{i+1} are distinct. The set of all such reduced expressions, with multiplication defined in the natural way, is the description of the *free group* $F(S)$ given in

Corollary I.12.6.) In the second case, the type of ''handle'' one can generally get is a *classification* of elements of U based on their behavior under the maps. Here it is the set of ''classes'' in this classification that turns out, in favorable circumstances, to form a universal object, i.e., a terminal object of the auxiliary category. (For instance, given an ''unknown'' group K with homomorphisms α, β into ''known'' groups G and H, we may ''classify'' elements $k \in K$ by the pairs of elements $\alpha(k) \in G$ and $\beta(k) \in H$. Given two such pairs $(\alpha(k), \beta(k))$ and $(\alpha(k'), \beta(k'))$, it is easy to write down the formula for the pair associated with kk', and this formula can be used to put a group operation on the set of all pairs consisting of an element of G and an element of H. The resulting group, $G \times H$, is easily shown to be the terminal object in the category of groups given with maps into G and H.) It is perhaps best to think about these abstract heuristics *after* one has digested a sufficient number of particular cases. How many are ''a sufficient number'' will depend on the student.

Notes on free groups and coproducts.

Is every subgroup of a free group free? Yes, but the proof is intricate (in contrast to the proof for free abelian groups), and we will not give it here. It can be found in the book of Magnus, Karrass and Solitar mentioned above. (Curiously, one useful way to visualize the result is topological. For the student with some familiarity with these matters: One shows that the fundamental group of any 1-complex is a free group, and that every free group occurs as such a fundamental group; one proves that subgroups of fundamental groups correspond to fundamental groups of covering spaces; and finally, one notes that a covering space of a 1-complex is a 1-complex.) Exercise I.12:1 shows that a free group on 2 generators can have a subgroup which is free on infinitely many generators, which is very different from the situation for free abelian groups (but again easily picturable in terms of covering spaces).

Why does Lang give an abstract proof of the existence of coproducts of groups in Proposition I.12.3 when he will later construct them explicitly? The abstract proof is extremely versatile; once you have digested it (or another such general proof), you can see instantly, on encountering any of a wide range of sorts of categories, that these have coproducts of arbitrary families of objects. The *explicit* description is particular to groups. When one encounters a new sort of algebraic object, and realizes by general arguments that these have coproducts, the next step is to ask whether there is, similarly, some elegant description of the *structures* of such coproducts.

Products and coproducts of groups presented by generators and relations.

Given two groups

$$G = \langle X \mid R \rangle, \qquad H = \langle Y \mid S \rangle,$$

with X and Y disjoint, it is straightforward to show by universal properties that the coproduct of G and H has the presentation $\langle X \cup Y \mid R \cup S \rangle$. The (direct) product, on the other hand, may be presented as $\langle X \cup Y \mid R \cup S \cup C(X, Y) \rangle$, where $C(X, Y)$ denotes the set of relations $xy = yx$ $(x \in X, y \in Y)$.

Some morphisms of functors.

As I mentioned in my comments to p.62 of Lang (and as Lang notes on p.63), one example of a functor **Set** $\to$ **Ab** is the construction associating to every set S the free abelian group $\mathbf{Z}S$. (Elsewhere, the group $\mathbf{Z}S$ has been characterized either as initial in a certain auxiliary category, or as the representing object for the functor **Ab** $\to$ **Set** taking an abelian group A to the set of all S-tuples of elements of A. The way in which this family of objects that were constructed separately as representing a family of functors, equivalently, as initial in a family of related auxiliary categories, can be ''tied together'' into one functor, is a case of a concept called *adjoint functors*, which I will sketch later in my comments on §II.3. However, without knowing anything about that general concept, you should not find it hard, using the universal properties of free abelian groups, to construct the morphisms among these groups needed to make the free abelian group construction a functor.) Similarly, the constructions of the *free group* on a set, and the *abelianization* G/G^{c} of a general group G, can each be made into a functor in a natural way. Let

us give ad hoc names to these and one other functor:

F: **Set** $\to$ **Group** for the free group functor,

F_{ab}: **Set** $\to$ **Ab** for the free abelian group functor,

A: **Group** $\to$ **Ab** for the abelianization functor, and

I: **Ab** $\to$ **Group** for the inclusion functor (defined in the obvious way).

With this list of functors to work with, I can now give some more examples of morphisms and isomorphisms of functors. Namely, suppose we associate to each set X the unique homomorphism $h(X) : F(X) \to F_{\mathrm{ab}}(X)$ sending the canonical image in $F(X)$ of every $x \in X$ to the image of the same element in $F_{\mathrm{ab}}(X)$. It is easy to verify that these maps together comprise a morphism of functors $f : F \to IF_{\mathrm{ab}}$. You should likewise not find it hard to show that for every set X, there is a canonical isomorphism $i(X) : AF(X) \cong F_{\mathrm{ab}}(X)$, and that these constitute an isomorphism of functors $i : AF \cong F_{\mathrm{ab}}$. If you'd like two more examples to think about, you should be able to find an isomorphism between the composite AI and the identity functor of the category **Ab**, and a certain natural morphism from the identity functor of **Group** to the reverse composite, IA.

Chapter II. Rings.

Re §II.1. Rings and homomorphisms.

P.83, definition of a ring [>]:

Just as the fundamental examples of groups are groups of permutations of sets, so there is a fundamental class of examples of rings: the rings of endomorphisms of abelian groups.

Indeed, recall that one can add and subtract members of $\mathrm{Hom}(A, B)$ for any abelian groups A and B, making this hom-set an abelian group. Taking $A = B$, this gives us an abelian group structure on $\mathrm{End}(A) = \mathrm{Hom}(A, A)$. Moreover, one can compose such endomorphisms, getting a ''multiplication'' on this set, for which the identity endomorphism will be a neutral element; and it is straightforward to verify that these additive and multiplicative structures together make $\mathrm{End}(A)$ a ring.

Just as every group G can be represented faithfully by permutations of a set, namely, via its action by left translations on its own underlying set, so a ring R can be represented faithfully by endomorphisms of an abelian group, namely, via its action by left multiplication on its own underlying additive group.

Note that the distributive law states that multiplication is a *bilinear* map with respect to the additive group structure.

P.85, **the convolution product** [~]: This is an important class of examples, but I will postpone discussion until Lang returns to it at greater length, two sections from now (on p.104). (The background of the term ''convolution product'' will get a discussion in itself.)

Note that in the sentence containing the last display, Lang modifies his viewpoint in two ways: First, by regarding the formal sums of the preceding display as functions on G, with a corresponding change of notation; and second, by not insisting that all but finitely many $f(x)$ be zero; though as he notes in the next sentences, *some* condition is needed to make the summations finite, and the most obvious such condition is the finiteness one. (But there are others. E.g., in the exercise he refers to, if we think of the monoid of positive integers under multiplication as a submonoid of the group of positive rationals, then the condition that only members of that submonoid have nonzero coefficient makes the multiplication work.)

Here is an interesting example of a ring of endomorphisms of an abelian group. Let A be the additive group of all polynomials in one indeterminate x with real coefficients. (Though we have not yet read the section on polynomials, in this and some subsequent examples I will assume familiarity with elementary properties of polynomials over the reals.) Let $X\colon A \to A$ denote the endomorphism of this abelian group

given by multiplication by x, let $Y\colon A \to A$ denote the endomorphism given by *differentiation* with respect to x, and for each real number r, let the same symbol r denote the endomorphism of scalar multiplication by r. The ring R that these operators generate is called the *Weyl algebra* over the real numbers. From the product law for differentiation, one gets the formula

(c15) $$YX - XY = 1.$$

Using this formula, one can reduce any member of R to the form $\Sigma\, r_{ij} X^i Y^j$ with real coefficients r_{ij}. These expressions look like ordinary polynomials in two variables, but the multiplication, based on (c15), is clearly different. (This algebra is important in quantum mechanics, where multiplying the wave function ψ of a particle by a coordinate function x corresponds to measuring the x-coordinate of the *position* of the particle, differentiating ψ with respect to x corresponds to measuring the x-coordinate of the particle's *momentum*, and the noncommutativity of these operators somehow corresponds to the impossibility of measuring these two quantities simultaneously!)

P.86, middle [=]: Change the sentence containing the word **principal** to ''A left ideal is called **principal** if it can be generated by one element, i.e., has the form Aa for some $a \in A$.''

Later in the same paragraph, Lang introduces the notation $(a_1, \dots, a_n)$, which he says is used for the *left* ideal generated by $a_1, \dots, a_n$. Actually, this notation is rarely used except in commutative ring theory, where left, right, and two-sided ideals are the same. Moreover, I generally prefer the more explicit notation $a_1 A + \dots + a_n A$ (except in cases where the symbol for the ring is complicated!) With that explicit notation, one can distinguish in the noncommutative case between the right ideal $a_1 A + \dots + a_n A$, the left ideal $A a_1 + \dots + A a_n$, and the two-sided ideal $A a_1 A + \dots + A a_n A$.

Same page, 5 lines from bottom [~]: Where Lang writes ''a **principal** ring'', he is translating the French term; the standard term in English, which we will use, is ''principal ideal ring''.

After the above line [>]: In the remainder of this section, Lang will a couple of times use the term ''prime ideal'', which he doesn't define till the beginning of the next section, and whose definition depends on a term defined at the end of this section. So turn now to the paragraph beginning on the bottom of p.91, and read through the first paragraph of §II.2 on p.92, where ''prime ideal'' is defined, then return to the next paragraph of this Companion... .

Having told you to read Lang's definition of ''entire ring'' on p.91, I must now say that this term is simply not used. As Lang notes, the standard term in English is ''integral domain'', and we will use it. Where Lang feels the need for an adjective, as in ''R is entire'', one says ''R is an integral domain''. ''Integral domain'' is often shortened to ''domain'', especially in longer phrases, so that what Lang calls a ''principal entire ring'', i.e., a principal ideal ring that is an integral domain, is commonly called a ''principal ideal domain''. This is an important concept, often abbreviated ''PID''.

Now return to p.86 of Lang, bottom paragraph.

P.87, third and fourth paragraphs [<]: Skip these unless you are familiar with the topics (rings of algebraic integers in number fields, and the ring theory of analytic functions on the complex plane).

P.88, top two paragraphs [<]: I recommend skipping these.

P.88, **Example** occupying two long paragraphs [<]: This is another interesting topic which would have to be treated at greater length to get anything out of it. So read it, skim it, or skip it as you like.

P.90, first line of paragraph containing first display [=]: For ''commuting with'' read ''centralizing''. (A set X is said to *centralize* a set Y if every element of X commutes with every element of Y. A statement that a set X ''commutes with'' Y is vague; it might be taken to mean $XY = YX$ as sets of products.) Note further that although the concept of adjoining to a ring A a set S of elements centralizing A is of interest in ring theory, one should definitely not, as Lang does, limit the definition of ''ring generators for B over A'' to the situation where S centralizes A! Finally, note that noncommutative ring theorists generally write the ring generated over a subring A by a set of elements S

as $A\langle S\rangle$.

P.90, line after first display [=]: Change ''n-tuples $(i_1, \dots, i_n)$ of integers ≥ 0'' to ''strings $(i_1, \dots, i_n)$ where n and $i_1, \dots, i_n$ are nonnegative integers''. (In other words, n is not assumed the same in all terms of the sum.)

Since the string $s_1, \dots, s_n$ can be different in different summands, the notation $a_{i_1 \dots i_n}$, which only shows the exponents and not the s's, is poor. In fact, since we must allow repetitions among the s_i, there is no need for exponents. So a better form for the display would be $\Sigma\, a_{s_1,\dots,s_n}\, s_1 \dots s_n$.

P.90, last sentence before Example [=]: By ''may not'' he means ''need not''.

P.91, three lines from bottom, definition of ''entire ring'' [=]: See my comment above.

P.92, end of §II.1 [>]. One good feature of Lang relative to Hungerford is that he does *not* introduce rings without unit. These are of some use in ring theory, but their place is secondary. Rather than try to develop the fundamentals of the theory of rings with unit in the context of the ''more general'' theory of rings without unit, it is most convenient, when basic facts on rings without unit are needed, to use a certain trick which reduces the study of rings without unit to that of rings with unit. We will not take the time to discuss this trick here; if you need it some day, ask me!

Re §II.2. Commutative rings.

P.92, beginning of the section [=]: Observe the proviso in italics! The results of this section are *not* in general true for noncommutative rings.

P.92, italicized result, *Every maximal ideal is prime* [~]: This does not need a separate proof: it will follow immediately from the result on the next page, that if $\mathbf{m}$ is maximal, $A/\mathbf{m}$ is a field, and the observation that a field has no zero-divisors.

(By the way, the text-formatting system that I am using doesn't have fraktur type, so I will use boldface lower-case letters $\mathbf{m}$, $\mathbf{p}$, $\mathbf{a}$, $\mathbf{b}$ etc. where Lang uses lower case fraktur letters for ideals.)

P.93, proof of first italicized statement [>]: In Exercise A2.2:1 it is shown that Zorn's Lemma can be used to obtain maximal proper subgroups of *finitely generated* groups, but not in general of infinitely generated groups. Why, then, can it be used here to get maximal ideals of not necessarily finitely generated rings? Because any ring A, even if not finitely generated as a *ring*, is finitely generated as an *ideal*, namely by $\{1\}$.

P.94, first display [=]: Lang says that the ring $\mathbf{Z}/n\mathbf{Z}$ will be abbreviated $\mathbf{Z}(n)$. This is very nonstandard; it is generally either written Z_n (like the group $\mathbf{Z}/n\mathbf{Z}$), or written out in full, as $\mathbf{Z}/n\mathbf{Z}$. In fact, Lang will rarely if ever use this notation he has introduced.

P.94, proof of the Chinese Remainder Theorem [~]: The proof Lang gives is messy. The easy-to-remember proof is as follows. Let A' denote the image of A in $\Pi_{i=1,\dots,n}\, A/\mathbf{a}_i$. For any pair of indices $i \neq j$, the hypothesis $\mathbf{a}_i + \mathbf{a}_j = A$ means that we can write $1 = u + v$, where u is an element of A whose image in A' has component 0 in the ith position, and v an element whose image has component 0 in the jth position. Thus, the image of v in A' has 1 in the ith position and 0 in the jth. Now fixing i, and multiplying together the images in A' of the ''v''s we get in this way for each j, we get an element having 1 in the ith position, and 0 everywhere else. Multiplying by an arbitrary element of A', we get an element having an arbitrary member of $A/\mathbf{a}_i$ in the ith position and 0 everywhere else. By summing a family of such elements, one chosen for each i, we can get an element of A' with arbitrary component in each place, proving that A' is all of $\Pi\, A/\mathbf{a}_i$, which is the content of the theorem.

P.95, top paragraph [~]: The proof of the asserted result can be called ''trivial'', but it is not obvious if you haven't seen it before. Multiply the given equation $\mathbf{a}_1 + \dots + \mathbf{a}_n = A$ by itself $\nu_1 + \dots + \nu_n$ times. (Actually, $(\nu_1 - 1) + \dots + (\nu_n - 1) + 1$ times is enough.) On the left-hand side of the resulting equation, note that every term contains *either* ν_1 factors $\mathbf{a}_1$, *or* ν_2 factors $\mathbf{a}_2$, *or* ν_3 factors $\mathbf{a}_3$, etc.. Hence that left-hand side is contained in $\mathbf{a}_1^{\nu_1} + \dots + \mathbf{a}_n^{\nu_n}$. Hence, since that left-hand side equals the right-hand side,

A, so does $\mathbf{a}_1^{v_1} + \ldots + \mathbf{a}_n^{v_n}$.

P.96, first two lines [~]: One can prove this result more easily by counting integers relatively prime to p^r.

P.96, statement of Theorem II.2.3 [~]: Here Lang is being sloppy (as he would say, he is ''abusing notation''), since he sometimes treats k as an element of $\mathbf{Z}/n\mathbf{Z}$ and other times as a member of $\mathbf{Z}$. I point this out because I don't want you to take such sloppiness as a model! The same applies to p.97, lines 3 and 4, and to the last line of the section.

In connection with the last line of the section, it intuitively makes sense to exponentiate an element of exponent n by a member of $\mathbf{Z}/n\mathbf{Z}$; but if you want to use this observation in a formal argument, you should explicitly state a convention, ''If x is an element of exponent n in a group G, and k is an element of $\mathbf{Z}/n\mathbf{Z}$, then we shall understand x^k to mean x^c, where $c \in \mathbf{Z}$ is any representative of the residue class k,'' and observe that this is well-defined, and that the ''laws of exponents'' hold for these generalized exponents.

It's not clear why Lang calls Theorem II.2.3 an ''application''. Perhaps he meant to apply the Chinese Remainder Theorem to express the rings $\mathbf{Z}/n\mathbf{Z}$ as direct products of rings $\mathbf{Z}/p^r\mathbf{Z}$, but forgot to go on to this point.

P.97, end of §II.2 [>].

More examples of commutative rings.

An important class of examples of commutative rings is given by factor rings of polynomial rings. Let us regard the ring $\mathbf{R}[X, Y]$ of polynomials in two variables over the real numbers as consisting of certain functions on the plane. Then if we restrict these functions to the circle $X^2 + Y^2 = 1$, we get the ring A of those real-valued functions on the circle that are given by polynomials in the two coordinate-functions X and Y. The restriction map $\mathbf{R}[X, Y] \to A$ is a surjective ring homomorphism, hence A can be identified with the factor-ring $\mathbf{R}[X, Y]/I$ for some ideal I. This ideal turns out to be the principal ideal generated by $X^2 + Y^2 - 1$.

In general, for an ideal I of $\mathbf{R}[X, Y]$, let us think of the factor-ring $\mathbf{R}[X, Y]/I$ as consisting of polynomial functions which are ''missing some information''. We shall see below that it may or may not be possible to express these, as in the above case, as restrictions of polynomial functions to subsets of the plane.

One ring-theoretic phenomenon with which we have had little experience is that of zero-divisors, and the various forms they take, so let us look at some examples of rings $\mathbf{R}[X, Y]/I$ with zero-divisors.

The ring $\mathbf{R}[X, Y]/(XY)$ can be identified with the ring of polynomial functions on the union of the X-axis and the Y-axis. (To show this, one must show that the kernel of the restriction map $\mathbf{R}[X, Y] \to$ **{**functions on the union of the axes**}** is precisely (XY). ''$\supseteq$'' is immediate, and ''$\subseteq$'' is an easy argument.) It has zero-divisors, because X and Y are nonzero elements with product 0. (Where I am writing X and Y, I mean, strictly speaking, the residue classes of these elements of $\mathbf{R}[X, Y]$ in this factor-ring. I am using the same symbols for elements of the polynomial ring and of the factor-ring for notational simplicity.) One can get another sort of description of this ring: Note that a function on the union of the axes is determined by its restrictions to each of the axes, which are a polynomial in X and a polynomial in Y respectively. It is not hard to show that a pair of polynomials $f(X)$ and $g(Y)$ can arise in this way if and only if $f(0) = g(0)$. Hence our ring can be described, up to isomorphism, as the subring of $\mathbf{R}[X] \times \mathbf{R}[Y]$ consisting of all pairs (f, g) such that $f(0) = g(0)$. (The zero-divisors noted above are $(X, 0) \cdot (0, Y) = (0, 0)$ in this description.) The use of two variable-symbols X and Y is just a holdover from the original picture; we can equally well identify our ring with the subring of $\mathbf{R}[X] \times \mathbf{R}[X]$ consisting of pairs satisfying $f(0) = g(0)$. This is, incidentally, an example of what is called a *subdirect product* (not to be confused with a *semidirect product*), meaning a subobject of a direct product, whose projection maps onto the two factors are both surjective.

We note a few more examples similar to the above. $\mathbf{R}[X, Y]/(Y(Y-1))$ can be identified with the ring

of polynomial functions restricted to the union of the two parallel lines $Y = 0$ and $Y = 1$. The restrictions to those two lines can be two arbitrary polynomials in X (verify this!), so this ring is isomorphic to the full direct product $\mathbf{R}[X] \times \mathbf{R}[X]$. $\mathbf{R}[X, Y]/(Y(Y-X^2))$ can be identified with polynomial functions restricted to the union of the line $Y = 0$ and the parabola $Y = X^2$. Restricting to these curves separately, we note that any polynomial function on either can be expressed as a polynomial in the X-coordinate alone, so our ring is again a subdirect product of two copies of $\mathbf{R}[X]$; in this case, it can be shown to consist of all pairs (f, g) whose values *and* first derivatives agree at 0. As a last example of this sort, $\mathbf{R}[X, Y]/((Y-X)Y(Y+X))$, the ring of polynomial functions restricted to the union of the three lines $Y = X$, $Y = 0$ and $Y = -X$, can be identified with the ring of 3-tuples of polynomials in X, (f, g, h), such that $f(0) = g(0) = h(0)$, and $f'(0) - 2g'(0) + h'(0) = 0$. (Proving this takes some thought, but still turns out to be a straightforward computation.)

We run into a new phenomenon when we look at $\mathbf{R}[X, Y]/(X^2)$. The subset of the plane defined by the equation $X^2 = 0$ is simply the line $X = 0$. But when we map $\mathbf{R}[X, Y]$ by restriction to functions on this line, the kernel consists of all multiples of X, which is larger than the ideal (X^2). Hence the ring of functions on the line induced by polynomials on the plane is not isomorphic to the ring we want to look at.

The trouble is that when we restrict to the line, we ''forget'' too much. The image in $\mathbf{R}[X, Y]/(X^2)$ of a polynomial $f(X, Y)$ ''remembers'' not only the terms without X, but also the terms *linear* in X. These terms specify the *partial derivative* of f in the normal direction to the line $X = 0$, all along that line. Algebraic geometers picture $\mathbf{R}[X, Y]/(X^2)$ as the ring of polynomial functions on an infinitesimally ''thickened'' version of the line $X = 0$, where there is just enough ''thickening'' so that the restriction of a function to it determines the normal derivative of that function, but not higher derivatives.

That $\mathbf{R}[X, Y]/(X^2)$ has zero-divisors is shown by the relation $X^2 = 0$. In terms of our ''values and derivatives'' interpretation, this is an instance of the observation that two differentiable functions whose values are zero all along the line $X = 0$ have product whose value *and* normal derivative are zero along that line.

A variant of this example is $\mathbf{R}[X, Y]/(X^2, XY)$. An element of this ring specifies the values of a polynomial function along the line $X = 0$, and its normal derivative to that line at the point $(0, 0)$ only, so it is a line ''thickened'' only at that point.

I have given examples in terms of polynomials over the real numbers because these are familiar and easy to visualize; but the relation between ideals of this ring and subsets of $\mathbf{R}^2$ has weaknesses; e.g., ideals such as $(X^2 + 1)$ or $(X^2 + Y^2 + 1)$ cannot be characterized by their behavior on $\mathbf{R}^2$. One gets much better behavior by looking at ideals in polynomial rings over the complex numbers $\mathbf{C}$; or for that matter by looking at ideals in polynomial rings over the real numbers, rational numbers, etc., in terms of the induced functions on space $\mathbf{C}^n$.

The study of ideals in the ring of polynomials in n indeterminates (equivalently, of homomorphic images of this ring), and the subsets of $\mathbf{C}^n$ that these determine, is the starting point of algebraic geometry. Lang proves in §IX.2 (which we do not read in 250A) some results on zero-sets of polynomials, and in §IX.5 he introduces the abstract object, the ''prime spectrum'' of a ring, which modern algebraic geometers study in place of these polynomially defined sets. But he does not make explicit the relationship between these spectra and the classical polynomially defined subsets, nor does any standard introduction to the subject that I have seen. I discussed that relationship in an expository article, *The Zariski topology and generalizations*, which appeared in a somewhat obscure volume (*Topology and its Applications*, the proceedings of a symposium in Budva, Jugoslavia, August 1972, published in Beograd in 1973). I can give anyone interested a reprint. (The article assumes a knowledge of elementary point-set topology.)

Re §II.3. Polynomials and group rings.

Before reading this section of Lang, I recommend reading sections III.1-III.4, and the notes on those

sections in this Companion. I will often assume familiarity with that material below.

P.99, first line [=]: Lang says the elements of the free generating set of a polynomial ring will be called ''variables'' (and makes a similar remark on p.104). This is an old-fashioned and misleading term. Although the concept of polynomial ring is motivated by that of a ring of polynomial *functions* of a real or complex variable, in the abstract algebraic definition, nothing is ''varying''. The preferred term is ''indeterminate''.

P.99, middle, definition of **reduction map** [>]: The term is motivated by the case where φ takes $\mathbf{Z}[X]$ to $\mathbf{Z}_n[X]$ for some n (usually prime) by ''reducing coefficients modulo n''. Though the term is not as intuitive in other cases, e.g., when φ is the embedding of the integers in the rational numbers, Lang makes the general definition by analogy with the motivating case.

P.102, second paragraph [>]: I claim that what Lang has shown here is essentially that the polynomial ring $A[X_1, \dots, X_n]$ is the *free commutative A-algebra on n generators.* Indeed, given a commutative A-algebra B, and elements $x_1, \dots, x_n \in B$, let us show that there is a unique A-algebra homomorphism $A[X_1, \dots, X_n] \to B$ taking each X_i to x_i. To fit our presentation to Lang's, let us first consider the case where A is a subring of B, and the homomorphism $A \to B$ defining the A-algebra structure of B is the inclusion. Then what Lang calls the *evaluation map* determined by $x_1, \dots, x_n \in B$ will be the unique A-algebra homomorphism with the desired property.

In the general case, if we let A' denote the image of A under the homomorphism that makes B an A-algebra, then by combining the ''reduction map'' $A[X_1, \dots, X_n] \to A'[X_1, \dots, X_n]$ with the ''evaluation map'' $A'[X_1, \dots, X_n] \to B$ determined by $x_1, \dots, x_n$, we get the desired unique homomorphism. But there is really no need to divide the construction of this homomorphism into two stages as Lang has done; if $a \mapsto \bar{a}$ is the homomorphism defining the A-algebra structure of B, one can define the homomorphism required by the universal property (for which ''substitution map'' is still an appropriate term) in one step, as $\Sigma\, a_{(\nu)}\, X_1^{\nu_1} \dots X_n^{\nu_n} \;\mapsto\; \Sigma\, \bar{a}_{(\nu)}\, x_1^{\nu_1} \dots x_n^{\nu_n}$.

P.102, last paragraph, definition of **occur** [>]: Lang here gives the word ''occur'' a technical meaning, to allow him to conveniently refer to the set of monomials which have nonzero coefficient in f. A more standard way of doing this is to call that set, $\{M_{(\nu)}(X) \mid a_{(\nu)} \neq 0\}$, the *support* of the polynomial f.

P.104, beginning of material on group and monoid rings, through end of §II.3 on p.107 [~]:

The idea of a group (*or monoid*) *ring.*

Suppose A is a commutative ring, and we are interested in an action of a group or monoid G by module endomorphisms on an A-module M. In this situation, note that the underlying additive group of M has two systems of endomorphisms given: the actions of the elements of A, and the actions of the elements of G. Now on the one hand, these systems of abelian group endomorphisms are subject to the usual conditions defining an A-module structure, respectively a G-set structure; on the other hand, the statement that G acts by *A-module endomorphisms* means that the action of each element of G *respects* the operations of the elements of A; in other words, the two sorts of operations commute with one another: $gax = agx$ ($g \in G$, $a \in A$, $x \in M$). We may bring these systems of operations together by looking at the ring of additive-group endomorphisms of M that they generate. We find that we can write elements of this ring as linear combinations of operations induced by members of G, with coefficients taken from A, and that composition of such linear combinations is described by the law given on p.104 (and also p.85).

This suggests that we construct a ring whose elements are formal linear combinations of elements of G with coefficients in A, and whose multiplication is given by that rule. We find that this construction indeed gives a ring AG, and we verify that a left AG-module is *equivalent to* an A-module with an action of G on it by A-module endomorphisms!

Let us be more precise about how AG is defined: to pin down the concept of ''formal linear combinations of elements of G with coefficients in A'', we observe that such an element should be

determined by the coefficients of the elements of G, and that all but finitely many of these coefficients should be zero. That is, such an element will be specified by a function associating to each element of G an element of A, such that the element associated to almost all elements of G is 0. Hence, the set of such functions is taken as the formal definition of the underlying set of AG (Lang, p.104). We then introduce the convenient abbreviation of using the symbol for an element of G to denote the image of that element in AG; formally the function which has value 1 on that element and 0 elsewhere, so that all elements can be written $\Sigma_{x\in G}\, a_x x$. (We are assuming G written multiplicatively. In the contrary case, we have to do some adjusting. E.g., to describe $A\mathbf{Z}$, we take a group written multiplicatively that is isomorphic to $\mathbf{Z}$, say $\langle x\rangle = \{x^i \mid i\in\mathbf{Z}\}$, then let our ring consist of formal linear combinations $\Sigma\, a_i x^i$ $(a_i\in A,\ i\in\mathbf{Z})$.)

Actually, this rigamarole of ''almost everywhere zero functions from G to A'' is just a way of describing the *free A-module* on the set of elements of G, which Lang cannot refer to directly at this point because he hasn't introduced modules! So the easier description of the underlying set of the monoid ring is as the free A-module on the underlying set of G.

Once the underlying set of this ring has been described as a free A-module, the natural way to describe its multiplication is to say that it is the unique A-bilinear map $\beta\colon AG\times AG\to AG$ which, on the basis G, agrees with the multiplication of G. The existence of such a bilinear map follows from Lemma III.4.c1. To see that it is associative, note that the ternary operations whose values on (u,v,w) are $\beta(u,\beta(v,w))$ and $\beta(\beta(u,v),w)$, respectively, are trilinear, and agree on elements of $X\times X\times X$; hence by the uniqueness part of that lemma, they are equal.

Incidentally, where Lang writes $A[G]$, I have substituted the symbol AG, preferred by group theorists. The general principle is that if X is a set, then a symbol of the form $A[X]$, since it resembles the symbol for a polynomial ring, is used to denote rings of linear combinations of *products* of elements of X with coefficients in A, while AX is used for rings or modules of linear combinations of elements of X alone with coefficients in A. The monoid algebra construction is of the latter sort. (Actually, there is a further distinction made. Suppose, for concreteness, that X consists of two elements, x, y. Then $A[X]$ is generally used for rings whose elements can be written as linear combinations of terms $x^i y^j$, while if one does not have enough commutativity or similar conditions to reduce to such a form, but needs to include separately terms such as x^2y, xyx and yx^2, one generally writes the algebra as $A\langle X\rangle$ or $A\langle x,y\rangle$.)

A terminological point: the ring AG should logically be called a group or monoid *algebra*. Both the terms ''group (or monoid) ring'' and ''group (or monoid) algebra'' are used in the literature. Lang, of course, cannot say ''algebra'' here, because he won't define that concept till Chapter III.

Some easily verified but important properties of monoid algebras are given in the following

Lemma II.3.c1. *Let A be a commutative ring. Then*

(i) *For any monoid G, the monoid algebra AG has the universal property that given an A-algebra R and a homomorphism h of G into the multiplicative structure of R, there exists a unique A-algebra homomorphism $AG\to R$ extending h.* (*I.e., AG is initial in the category of A-algebras given with homomorphisms of G into their multiplicative structures.*)

(ii) *For any set X, the monoid algebra on the free abelian monoid on X is the free commutative A-algebra on X; i.e., the polynomial algebra $A[X]$.*

(iii) *For any set X, the monoid algebra on the free monoid on X is the free A-algebra on X* (*generally written $A\langle X\rangle$*).

(iv) *An A-module M together with an action of a monoid G on M by A-module endomorphisms* (*with $1\in M$ understood to act by the identity*) *is equivalent to an AG-module, under the correspondence sketched earlier in this remark.*

Sketch of Proof. (i): Because AG is free as a module on the underlying set of G, the universal property of free modules allows us to extend the set-map h to an A-module-homomorphism $AG \to R$. If one now calls on the fact that $h\colon G \to R$ is assumed to respect multiplication, one can verify that A-linear combinations of elements $h(g)$ $(g \in G)$ multiply in the ring R by the same formula that defines the multiplication in the monoid algebra, and hence that the extension of h respects ring multiplication.

(ii): It will suffice to show that this monoid algebra has the universal property of the polynomial algebra. Given any map $f : X \to R$, where R is a commutative A-algebra, the universal property of the free abelian monoid G on X allows us to extend this in a unique manner to a homomorphism from that monoid to the multiplicative monoid of R. The universal property proved in (i) above for the monoid algebra now allows us to extend this map in a unique manner to an algebra homomorphism $AG \to R$, establishing the desired universal property.

(iii) is proved in the same way. Incidentally, the free monoid on X (respectively, the free abelian monoid on X) can be shown to be isomorphic to the sub*monoid* generated by X in the free group (respectively the free abelian group) on X.

(iv) is the result in terms of which we motivated the construction of AG! To formalize the argument, note that an action of G on an A-module M means a monoid homomorphism $G \to \mathrm{End}_A(M)$. By (i), this is equivalent to an A-algebra homomorphism $AG \to \mathrm{End}_A(M)$, which in particular gives a ring homomorphism from AG to the ring of abelian group endomorphisms of M, i.e., an AG-module structure on M. Conversely, every AG-module structure gives an A-module structure and an action by G respecting that structure. It is straightforward to verify that these ways of going between one and the other structure on M are inverse to one another. □

Note that statement (iv) of the above lemma means that linear actions of G on A-modules (in particular, on A-vector-spaces if A is a field) may be studied by looking at the structure of the algebra AG, and its modules. This is an important principle in the theory of groups and their representations. Chapter XVIII of Lang (not covered in 250A) gives an introduction to this technique in the case where G is a finite group.

There is a pair of longstanding open questions concerning group algebras:

Open Questions II.3.c2. *Let K be a field and G a torsion-free group, i.e., a group with no nonidentity elements of finite order. Must the following statements be true?*

(i) (*Zero-divisors conjecture.*) *KG is a ring without zero-divisors.*

(ii) (*Trivial units conjecture.*) *The only invertible elements of KG are the products cg $(c \in K - \{0\}$, $g \in G)$.*

When Lang first sketchily introduced group algebras as an Example on p.85, he called their multiplication operation ''the convolution product''. For those who are curious, let me describe the

Background of the term ''convolution'' (optional).

This term actually goes back to a very different line of thought.

Suppose f is a ''non-smooth'' function on the real line, such as the function which is 1 on $[0, 100]$ and 0 elsewhere, and we seek to ''smooth'' it out by ''blurring'' it. Intuitively, a nice way to do this would be if we could replace the ''value'' at each point by a bell-shaped curve, and integrate these together. In fact, if we write $g(x) = e^{-x^2}/\sqrt{\pi}$, we find that the above idea can be made precise, by taking for the ''blurred'' variant of f the function $f * g$ defined by

$$(f * g)(x) \;=\; \int_{-\infty}^{+\infty} f(x-y)\,g(y)\,dy. \tag{c16}$$

The idea is that the above definition makes the value of $f * g$ at x depend not only on the value of f at

x, but also on the value at nearby points $f(x-y)$, with the contributions of these points ''weighted'' by the smooth function $g(y)$.

The above is just one way of motivating the operation (c16). I hope you can see that in general, (c16) describes an interesting way of combining two functions f and g (satisfying appropriate conditions to make the integral converge) to get a third. The name given to this construction is ''convolution'', meaning ''twisting together''. I will mention without proof one other property of this construction: the Fourier transform of the convolution of two functions is the *product* of their Fourier transforms; and conversely, the Fourier transform of their product is the convolution of their Fourier transforms.

If we look at functions on the integers instead of on the real line, we may define an analog of (c16) by using a sum instead of an integral:

(c17) $$(f * g)(n) = \sum_{m=-\infty}^{+\infty} f(n-m)g(m).$$

As before, we need a condition to insure that these infinite sums will make sense. The easiest such condition is that the functions f and g be zero almost everywhere; this allows us to make sense of the above definition for functions with values in an arbitrary ring A, where we do not necessarily have a concept of convergence. But in that case f and g can be thought of as elements of the group algebra $A\mathbf{Z}$; and our ''convolution multiplication'' turns out to be the multiplication of the group algebra. (To see this, note that the arguments of f and g in (c17), $n-m$ and m, range over all pairs of integers summing to n.) Hence, by extension, the term ''convolution multiplication'' is often, as in Lang, used for the multiplication operation of any group algebra.

If in (c17) we restrict f and g to be functions on the *nonnegative* integers, modifying the range of summation appropriately, we of course get the description of multiplication of polynomials, looked at as ''sequences of coefficients''.

P.105, first line of text [=]: ''triples $(u\ v,\ y)$'' should be ''triples (u, v, y)''.

P.106, Proposition II.3.1 [~]: Observe that the condition that $h(a) = a$ for all $a \in A$ is equivalent to saying that we are dealing with a homomorphism of A-algebras (though Lang does not introduce that concept till the next Chapter).

P.107, end of §II.3 [>].

The concepts of this section have given us enough examples of rings so that we can give some further classes of

Examples of modules.

Let A be a commutative ring, and as in Lemma II.3.c1 let $A\langle x, y\rangle$ denote the free associative A-algebra on two generators x, y, equivalently, the monoid algebra over A on the free monoid $\langle x, y\rangle$ on two generators. What does a module over $A\langle x, y\rangle$ look like? Let us think of such a module as given by an A-algebra homomorphism $A\langle x, y\rangle \to \mathrm{End}_A(M)$ for some A-module M (cf. Lemma III.1.c3(b)). The universal property of the free algebra $A\langle x, y\rangle$ says that such a homomorphism is determined uniquely if we specify, in any way, elements of $\mathrm{End}_A(M)$ to which the free generators x and y should be mapped. So a module over $A\langle x, y\rangle$ is determined by choosing any A-module M, and any two A-module endomorphisms ξ and η of M, and letting multiplication by x act as ξ, and multiplication by y act as η. You can make up examples ad lib, taking, say, $A = \mathbf{Z}$, letting M be the free $\mathbf{Z}$-module (free abelian group) on n generators, using any two $n \times n$ matrices over $\mathbf{Z}$ to specify two endomorphisms, and letting $\mathbf{Z}\langle x, y\rangle$ act as just described.

Modules over the polynomial algebra $A[x, y]$ have an identical description, except that ξ and η, rather than being any two endomorphisms, must be two endomorphisms that commute with one another.

I chose the case of two indeterminates simply for concreteness; the analogous characterizations hold for any number of indeterminates. A particularly simple and important case is that of a single indeterminate. In that case we see that $A\langle x\rangle$ and $A[x]$ are the same; let us use the latter symbol. Thus, a module over

$A[x]$ is equivalent to an A-module together with a single endomorphism ξ. In particular, if A is a field, this is a vector-space given with a single linear operator. In §IV.1 we shall see that if A is a field, $A[x]$ is a principal ideal domain. Now the theory of finitely generated modules over a PID (Lang, §III.7, not covered in 250A) is quite similar to the theory of finitely generated $\mathbf{Z}$-modules (a.k.a. finitely generated abelian groups) that we saw developed in §I.8. This theory can be used to give an elegant analysis of the structure of a finite-dimensional vector spaces given with a single linear operator; equivalently, a theory of similarity invariants for matrices over a field; Lang will do this in Chapter XIV.

The next topic could have been introduced back when we were reading §I.11 or I.12, or could be put off to a later section than this; but let me introduce it here, taking the monoid algebra concept as a springboard.

Adjoint functors – a sketch.

The universal property of the monoid algebra AG involves considering homomorphisms from a *monoid* G to a *k-algebra* R (point (i) of the lemma on monoid algebras in this Companion). But if we look at what this means, we see that we are really considering homomorphisms from G to the *monoid* of elements of R under multiplication. Thus, if we write U for the ''forgetful functor'' from A-algebras to monoids that forgets the A-module structure of an algebra and remembers only the multiplication, we can say that we are dealing with homomorphisms $G \to U(R)$. This is more sensible, since both objects now lie in the category of monoids. The universal property of the monoid algebra then says that AG is a A-algebra given with a homomorphism $\eta_G\colon G \to U(AG)$, such that for any A-algebra R, and homomorphism $f : G \to U(R)$, there is a unique A-algebra homomorphism $h\colon AG \to R$ such that $f = U(h)\eta_G$.

You should now check for yourself that the universal properties of *free group*, *free abelian group*, *free commutative A-algebra*, and *Grothendieck group* all have analogous forms, gotten by substituting for ''the forgetful functor from A-algebras to monoids'' the forgetful functor from groups to sets, from abelian groups to sets, etc.. (The symbol ''U'' and its alphabetical neighbors are often used for these forgetful functors because many of them can be described as ''underlying set'' functors, or the like. Similarly, ''F'' and its alphabetical neighbors are often used for what we shall soon be calling the ''left adjoints'' of these functors, because of the prototypical case of *free* object constructions.)

Given a monoid homomorphism $\varphi\colon G \to H$, Lang has noted that we get a ring homomorphism $AG \to AH$. The existence of this homomorphism also follows from our universal property: the canonical homomorphism $\eta_H\colon H \to U(AH)$, composed with the given homomorphism $\varphi\colon G \to H$ yields a homomorphism $G \to U(AH)$; the universal property of AG now implies that this arises from a homomorphism $\bar{\varphi}\colon AG \to AH$; precisely, that there is a unique $\bar{\varphi}$ such that $\eta_H\, \varphi = U(\bar{\varphi})\eta_G$. It is not hard to show that the construction of $\bar{\varphi}$ from φ makes the monoid algebra construction a *functor*, which is again analogous to what we have seen is true for free groups, free rings, etc..

Let us state the phenomenon that we are observing in a general category-theoretic context:

Let $\mathbf{C}$ and $\mathbf{D}$ be any categories, and $U\colon \mathbf{C} \to \mathbf{D}$ a functor. Suppose that for every object D of $\mathbf{D}$, there exists an object $F(D)$ in $\mathbf{C}$, and a morphism $\eta_D\colon D \to U(F(D))$, with the universal property that for every object C of $\mathbf{C}$ and morphism $f : D \to U(C)$, there exists a unique morphism $g\colon F(D) \to C$ such that $f = U(g)\eta_D$. Then we can make F into a functor $\mathbf{D} \to \mathbf{C}$ in a natural way, and this is called the *left adjoint* to the given functor U. Likewise, U is called the *right adjoint* to F; the relation between U and F is called an *adjunction*. Given such an adjunction, one finds that for every pair of objects $C \in \mathrm{Ob}(\mathbf{C})$, $D \in \mathrm{Ob}(\mathbf{D})$, the relation between morphisms $f : D \to U(C)$ and $g\colon F(D) \to C$ described above is a *bijection*

$$\mathrm{Hom}_{\mathbf{D}}(D,\, U(C)) \;\longleftrightarrow\; \mathrm{Hom}_{\mathbf{C}}(F(D),\, C).$$

(The names ''right'' and ''left'' adjunction refer to the fact that U occurs in the *right* slot in one hom-set, and F in the *left* slot in the other.)

A result which you should not find hard to prove, once you have digested the above concept (but for which you are not responsible) is

Lemma II.3.c3. *If* $\mathbf{C} \xrightarrow{U} \mathbf{D} \xrightarrow{V} \mathbf{E}$ *are functors having left adjoints* $\mathbf{C} \xleftarrow{F} \mathbf{D} \xleftarrow{G} \mathbf{E}$, *then the composite functor* VU *has left adjoint* FG. □

For example, if we denote by $\mathbf{Comm}_A$ the category of commutative A-algebras, and by **AbMonoid** the category of abelian monoids, then we know that the forgetful functors $\mathbf{Comm}_A \to \mathbf{AbMonoid} \to \mathbf{Set}$ (where the first functor forgets the additive structure and remembers the multiplication, while the second forgets even that) have for left adjoints the *monoid algebra* and *free monoid* functors respectively, and that their composite, the forgetful functor from commutative A-algebras to sets, has for left adjoint the *polynomial algebra* functor. Thus, the above lemma shows why point (ii) of the lemma on monoid algebras proved in this Companion, saying that polynomial algebras are monoid algebras on free commutative monoids, had to be true. Looking at not-necessarily-commutative A-algebras and monoids, one likewise gets point (iii) of that lemma. You can discover some further results of this sort by applying the above lemma to the forgetful functors $\mathbf{Ab} \to \mathbf{Group} \to \mathbf{Set}$, and likewise $\mathbf{Ab} \to \mathbf{AbMonoid} \to \mathbf{Set}$.

Though the above examples of functors having left adjoints were all *forgetful* functors, there are many that do not fit this description. For instance, you might find it interesting to prove that for any positive integer n, the functor taking every commutative ring A to the group $\mathrm{SL}_n(A)$ of $n \times n$ matrices over A with determinant 1 has a left adjoint.

We will not be able to study the subject of adjunctions further in this course, aside from noting a few more examples. For further development, see any text on category theory, in particular, my 245 notes or Mac Lane's *Categories for the Working Mathematician*. One of the most striking results developed in the 245 notes is a theorem of Peter Freyd (in Chapter 9) characterizing precisely the functors among varieties of algebras that have left adjoints. (The term ''variety of algebras'', defined there, includes all of the categories of mathematical objects that have arisen so far in this course, in particular **Set**, **Monoid**, **Group**, **Ring**, the category $\mathbf{Comm}_A$ for A a fixed commutative ring, the category R-**Module** for R any ring, and the category G-**Set** for G a fixed group.)

Note on automorphisms of polynomial algebras (*optional; > Exercise II.3:1*).

For A a commutative ring, Exercise II.3:1(a)-(c) gives a simple description of the automorphism group of the polynomial algebra $A[X]$. However, the study of automorphisms of polynomial algebras in several indeterminates is more complicated. For instance, for every polynomial $f(X) \in A[X]$, the algebra $A[X, Y]$ has an automorphism acting by $X \mapsto X$, $Y \mapsto Y + f(X)$. (What is its inverse?) Likewise, for every $g \in A[Y]$ we get an automorphism $X \mapsto X + g(Y)$, $Y \mapsto Y$. Though it is easy to see how automorphisms of the first type compose with each other, and how those of the second type compose with each other, maps gotten by alternately applying automorphisms of the two different types can get arbitrarily complicated. Nevertheless, it has been proved that if A is a field, the group of A-algebra automorphisms of the polynomial algebra $A[X, Y]$ is generated by the automorphisms of these two sorts, together with those of the more trivial sort $X \mapsto aX$, $Y \mapsto Y$ ($a \in A - \{0\}$). But for polynomials in more than two indeterminates, it is not known even whether the analog of this statement is true. (The subgroup of the automorphism group of $A[X_1, \ldots, X_n]$ generated by the automorphisms of these types is called the group of *tame automorphisms*, so the standard formulation of this question is, ''Is every automorphism of a polynomial algebra over a field tame?'' Incidentally, in the 2-variable case, though the automorphisms are, in the above sense, ''known'', there is another longstanding open question concerning these, the *Jacobian conjecture*, which I will mention at the end of the comments concerning §IV.1.)

We remark that when A is *not* a field, part (d) of Exercise II.3:1 can fail. Indeed, let k be any integral domain (say a field), and let $A = k[Y]$. Then the k-algebra automorphism of $k[X, Y] = A[X]$

that acts by $X \mapsto X$, $Y \mapsto Y + X$ is a ring automorphism of $A[X]$ which does not take A into itself.

Re §II.4. Localization.

P.108, construction of the ring $S^{-1}A$ [~]:

What is being constructed is a universal example of a ring A' with a homomorphism $u : A \to B$ which sends all elements of S to invertible elements of B.

The construction of this sort that most of you have probably seen is that of the field of fractions of an integral domain A (where $S = A - \{0\}$). To understand the details of the general construction, however, we must also think about the case where A has zero-divisors.

Suppose that in a commutative ring A one has a relation $sa = 0$, and that f is any homomorphism of A into a commutative ring A' such that $f(s)$ is invertible. The equation $sa = 0$ gives us $f(s)f(a) = 0$, and multiplying by $f(s)^{-1}$, we get $f(a) = 0$. This shows that if we wish to map a ring A to another ring in which every member of a multiplicative submonoid $S \subseteq A$ becomes invertible, then every element of A which is annihilated by a member of S must go to zero. Now suppose we want to test whether two elements, $f(a_1)f(s_1)^{-1}$ and $f(a_2)f(s_2)^{-1}$ are equal. By bringing them to the common denominator $f(s_1 s_2)$, we may write their difference $f(s_2 a_1 - s_1 a_2)f(s_1 s_2)^{-1}$. This will be zero if and only if $f(s_2 a_1 - s_1 a_2) = 0$, and by our preceding discussion, this will necessarily happen if $s_2 a_1 - s_1 a_2$ is annihilated by some element $s' \in S$.

Now a general principle in attempting to construct a mathematical object with certain properties is to collect all the conditions we can come up with on how it should look, and then try to build a model based on these conditions. If we have collected enough conditions, the model will ''work''; if not, we see how it fails to work, and try to extract from this further conditions. Having found that when a commutative ring A is mapped to another ring A', in such a way that elements of a given multiplicative submonoid S all become invertible, two elements of the form $f(a_1)f(s_1)^{-1}$ and $f(a_2)f(s_2)^{-1}$ will be equal in the latter ring if

(c18) $s_2 a_1 - s_1 a_2$ is annihilated by an element of S,

we can try constructing our ring to consist of formal expressions intended to represent elements $f(a)f(s)^{-1}$ (namely, ordered pairs (a, s)), and identify two expressions (a_1, s_1) and (a_2, s_2) precisely when (c18) holds. As noted in the reading in Lang, this ''works''. Since the arithmetic conditions we have imposed on ordered pairs (a, s) are conditions that must hold among the elements $f(a)f(s)^{-1}$ for *any* homomorphism f of A into a ring which maps elements of S to invertible elements, the map of A into the ring we have constructed is universal.

There are certain improvements one can make in Lang's verification that the above construction does indeed ''work''. For instance, because S is a multiplicative submonoid, it is easy to see that for any two equivalence classes u and v, there exist elements $(a, s) \in u$ and $(a', s) \in v$ with the same second component s; intuitively, that ''any two fractions can be written to a common denominator''. Now given $u, v \in S^{-1}A$, suppose we choose representatives, (a, s) and (a', s), with the same second component, and define $u + v$ to be the equivalence class of $(a + a', s)$. One must verify that this is well-defined (independent of our choice of expressions for u and v having common denominator). Once this is done, the verification that addition of fractions is *associative* is much more straightforward than when it is defined by a messy formula. One can also put the condition for two elements (a, s) and (a', s') to be equivalent in a form that is slightly more elegant and convenient to calculate with; namely that there exist $t, t' \in S$ such that $(ta, ts) = (t'a', t's')$.

The above construction is called ''localization'', but it will be easiest for me to explain why after Lang has introduced the concept of ''local ring'', later in this section.

Another way of approaching localization (sketch, optional).

How does a child who knows about the integers learn the meaning of ''1/2''? One of the things he or she learns is that certain integers can be ''halved''. In other words, ''1/2'' corresponds to a certain *partial map* from integers to integers: $0 \mapsto 0$, $2 \mapsto 1$, $4 \mapsto 2$, etc.. Similarly, $1/4$ corresponds to the partial map $0 \mapsto 0$, $4 \mapsto 1$, $8 \mapsto 2$, etc.. This suggests that $2/4$ should correspond to the partial map $0 \mapsto 0$, $4 \mapsto 2$, $8 \mapsto 4$, etc.. But we want to have $2/4 = 1/2$; so it seems that we should define an *equivalence relation* among partial maps, which relates maps which agree on a reasonable subset of the intersection of their domains.

When we abstract these ideas, we get the following: Let A be a commutative ring, and S a multiplicative submonoid in A. Let Φ denote the set of all *A-module homomorphisms* $sA \to A$ with $s \in S$, and let $\sim$ be the equivalence relation which puts $\varphi_1 : s_1 A \to A$ and $\varphi_2 : s_2 A \to A$ in the same equivalence class if and only if they have the same restriction to some ideal $s_3 A$ ($s_3 \in S$) contained in both $s_1 A$ and $s_2 A$. Let A' denote the quotient set $\Phi/\sim$. We can define ring operations on this set: we add two partial maps by restricting them to a common ideal and adding their restrictions; we multiply them by composing them, again after appropriate restriction. It is not hard to verify that we get, in this way, well-defined operations on equivalence classes of partial maps, and that these make A' a ring.

We may map A into this ring by sending each $a \in A$ to the equivalence class of the map ''multiplication by a'': $A \to A$. If the set $S \subseteq A$ consists of non-zero-divisors, it is easy to verify that the image of each $s \in S$ under the above homomorphism is invertible, with inverse given by the map $sA \to A$ taking sa to a ($a \in A$). In this situation, A' is, up to isomorphism, precisely the ring $S^{-1}A$ we are interested in. If some elements of S are zero-divisors in A, however, this does not work. (The above map $sa \mapsto a$ is not well-defined.) So to use this approach to localization, one must *first* divide A by the ideal I of all elements that are annihilated by members of S, then apply the above construction to A/I.

P.109, before third-from-last line [>]: We have just seen that under appropriate conditions, the map $A \to S^{-1}A$ is injective. There is a converse sort of result, saying that an injective ring homomorphism $A \to B$ with certain properties must be ''essentially'' this map $A \to S^{-1}A$.

Lemma II.4.c1. *Let A be an integral domain, and let S be a multiplicative subset which does not contain 0. Suppose $f : A \to B$ is a one-to-one ring homomorphism such that for every $s \in S$, $f(s)$ is invertible in B.*

Then the homomorphism $h : S^{-1}A \to B$ given by the universal property of $S^{-1}A$ is also one-to-one. Hence $S^{-1}A$ is isomorphic (via this map h) to the subring $\{ f(s)^{-1} f(a) \mid s \in S, a \in A \}$ of B.

Proof. Let a/s be a nonzero element of $S^{-1}A$; thus a is a nonzero element of A. As discussed by Lang on this page, $h(a/s) = f(s)^{-1} f(a)$. If this were zero, then multiplying by $f(s)$ we would get $f(a) = 0$, contradicting the assumption that f is one-to-one. Hence h has trivial kernel, so it is one-to-one, giving the first assertion. Hence h is an isomorphism between its domain, $S^{-1}A$, and its image, $\{h(a/s)\} = \{ f(s)^{-1} f(a) \}$, giving the second assertion. □

As Lang will note in point **2** on the next page, if we take S to be the set of all nonzero elements of A, then $S^{-1}A$ is what is known as the field of fractions of A. Hence the above Lemma shows that given any one-to-one homomorphism f of an integral domain A into a ring B in which all nonzero elements of A become invertible (for instance, any one-to-one homomorphism of A into a field B), the elements of the form $f(s)^{-1} f(a)$ in B form a subfield of B isomorphic to the field of fractions of A.

P.110, second line [=]: For ''In that case'' read ''For $S = A^*$''.

P.110, three lines before point ''**3.**'' [=]: The phrase ''From general localization properties'' is somewhat cryptic. I think what Lang means is that the localization of $A[X_1, \dots, X_n]$ obtained by inverting all polynomials which do not have the value zero at $(b_1, \dots, b_n)$ can be mapped homomorphically to K so as to send each X_i to b_i, by the universal property of localization (p.109), and, further, that this localization also embeds in $A(X_1, \dots, X_n)$, so that one can map the appropriate subring of that field homomorphically to K, by the indicated substitution.

P.110, first two lines of point ''**3.**'' [=]: In a general partially ordered set, a unique maximal element need not be a *greatest* element. However, the partially ordered set of proper ideals of a ring has the property that every element is $\leq$ a maximal element. (Do you remember why?) It follows that if a ring R has a unique maximal ideal, this is a *largest* proper ideal.

P.110, end of same paragraph [>]:

Background of the terms ''local ring'' and ''localization''.

Let us think of the ring of polynomials over the complex numbers, $A = \mathbf{C}[X]$, as a ring of functions on the complex plane. We know it is a principal ideal domain, and since every polynomial in this ring is a product of linear factors (something we will prove in a later chapter, but which I assume you have seen in an undergraduate course), the irreducibles are (up to units) simply the elements $X-c$ for $c \in \mathbf{C}$. Hence the general nonzero prime ideal of A is of the form $(X-c)A$. Note that this ideal can be described as consisting of all polynomials whose value at the point c is 0.

Since A is a unique factorization domain, we see that an element $f(X)$ of this ring will be divisible by an element $g(X)$ if and only if at each $c \in \mathbf{C}$ which is a zero of the polynomial g, the polynomial f has a zero of at least as high a multiplicity as g has. Within the field of fractions of A, regarded as the ring of ''rational functions'' on $\mathbf{C}$, the subring A can be characterized as consisting of the elements which do not ''blow up'' anywhere on the plane.

Now suppose we fix a point $c \in \mathbf{C}$, and blind ourselves to everything but the ''local'' picture in the immediate neighborhood of c. With our vision thus restricted, the best approximation of the condition ''f is divisible by g'' that we can perceive is ''the order of the zero of f at c is at least the order of the zero of g at c''. Thus, whenever two such elements f and g satisfy the above condition, we will accept f/g as though it were a member of our ring. The elements so ''accepted'' indeed form a ring, consisting of all rational functions which can be written with denominator not divisible by $X-c$; equivalently, which do not ''blow up'' at c. This is the ring called $A_{(X-c)A}$ (see Lang, p.110, middle, ''... and S^{-1} is denoted $A_{\mathbf{p}}$''). Since these rings arise by considering the ''local'' behavior of rings of functions, the type of rings they exemplify are called the *local rings*.

Although the construction of adjoining inverses to a set S of elements of a ring was *first* introduced to construct fields of fractions, it is in the above context that it required more careful study; hence the general construction of rings $S^{-1}A$ from rings A has come to be called ''localization''. (Note, however, that localization with respect to a multiplicative monoid does not in general yield a local ring.)

I described the construction of $\mathbf{C}[X]_{(X-c)}$ facetiously in terms of ''blinding oneself to everything but the local picture''. But in fact, it is not a pointless exercise. It turns out that many considerations of number theory and ring theory can be best studied by considering the behavior of a ring A ''one prime at a time'', and then putting these results together; and the local rings $A_{\mathbf{p}}$ are the most convenient tools for these ''one prime at a time'' considerations.

P.111, end of §II.4 [>].

Terminological warning. There is a certain amount of confusion between terms used for rings obtained by dividing a given ring A by an ideal, and by inverting a multiplicative set of elements. The first kind of ring, A/I, is sometimes called the ''factor ring'' and sometimes the ''quotient ring'' of A by the ideal I. The former term has the disadvantage that it can also mean one of the ''factors'' A_i in a direct

product decomposition $A = \prod A_i$ (though to avoid ambiguity the latter can be called a ''direct factor'' of A); the term ''quotient ring'' of course invites confusion with ''ring of quotients''. A localization $S^{-1}A$ may be called a ''ring of quotients'' or a ''ring of fractions''; the former term, as noted, invites confusion with one of the terms for the construction A/I. To avoid ambiguity, we shall follow Lang in calling the construction A/I a ''factor ring''. (We have likewise called a group G/N a ''factor group''). Unfortunately, in the case of an arbitrary set S and an equivalence relation $\sim$, the term ''quotient set'' is the only standard one.

The remaining notes on this section are optional.

''Specializations'' of fields (*sketch, optional*).

Recall that I translated Lang's remark in the two sentences before point **3** on p.110, as meaning that the set of rational functions in $X_1, \dots, X_n$ that could be evaluated at a point b of K^n forms a subring of the rational function field, which may be mapped homomorphically to K so as to send each X_i to b_i. Such *partial homomorphisms* from a field E to another field K – to be precise, homomorphisms φ from a subring $E_\varphi \subseteq E$ to K, having the property that if $x \in E_\varphi$ and $\varphi(x) \neq 0$, then $x^{-1} \in E_\varphi$ as well – are studied under the name *specializations* $E \to K$. (In this situation, it is easy to verify that the ring E_φ will be a *local* subring of E.) The concept has some surprises. For instance, suppose one has a specialization $\mathbf{Q}(X, Y) \to \mathbf{Q}$ taking X to 1 and Y to 2. This information does *not* completely determine φ, for on the element $(X-1)/(Y-2)$ it may be undefined, or it may be defined and take on *any* value! One can also have specializations from a field of characteristic 0 to one of finite characteristic, while this is clearly impossible for *homomorphisms* of fields.

One way of getting uniqueness is to study specializations which have minimal domain among all specializations whose domains contain a given set of elements (e.g., the set $\{X, Y\}$ in the above discussion). Using this idea, one can show that the rational function field $\mathbf{Q}(X_1, \dots, X_n)$ has a universal property with respect to specializations, which could justify calling it the ''free field'' on $X_1, \dots, X_n$, though there is no free field in the ordinary sense. Sometimes, *maximal* specializations are also studied; these are known as *places*. The concept of specialization is also used in studying division rings; cf. P. M. Cohn's books *Free Rings and Their Relations*, 2nd ed. 1985, and *Skew Field Constructions*, 1977; or the revised versions of these books, *Free Ideal Rings and Localization in General Rings*, 2006, and *Skew Fields*, 1995. (''Skew field'' is another term for ''division ring''. ''Specialization'', incidentally, is an older general term for ''substituting in particular values'', and Lang sometimes uses it in this sense.)

Localization of noncommutative rings (sketch, optional).

In noncommutative ring theory, there are several distinct versions of the concept of localization, with varying properties. In ''universal localization'', one takes a ring R and a set of elements $\Sigma \subseteq R$, and constructs, by generators and relations, a ring $R\langle \Sigma^{-1} \rangle$ in which images of elements of Σ are all invertible, and which is universal for this property. But the study of the structure of this ring is a much more nontrivial problem than in the commutative case. E.g., an element of the form $rs^{-1}t + uv^{-1}w$ cannot, in general, be reduced to an expression involving just a single ''denominator''. To get elements of more complicated forms, such as $(rs^{-1}t + uv^{-1}w)^{-1}$, it may be necessary to *repeat* the process of adjoining inverses more than once. Finally, even if R is without zero-divisors, it may be difficult to determine whether $R\langle \Sigma^{-1} \rangle$ is or isn't the trivial ring!

However, under certain special conditions, one can construct from a ring R and a subset S a ring $S^{-1}R$ in which every element has the same simple form $s^{-1}r$ as in the commutative case. One way to find the conditions under which this can be done is to generalize the ''partial homomorphisms'' approach to localization described above. A left R-module homomorphism from a left ideal Rs to R can be thought of as right multiplication by some element $s^{-1}r$; the conditions for the sum and composite of two

such maps to make sense turn out to be that for every $a \in R$ and $s \in S$, one have $Sa \cap Rs \neq \emptyset$. This is called the *left Ore condition* (after Oystein Ore), the construction of the ring $S^{-1}R$ in this case is called *Ore localization*, and when the set $S = R - \{0\}$ satisfies the left Ore condition, then R is called a *left Ore ring*, and $S^{-1}R$ turns out to be a division ring. Ore localization does not, in general, give rings with a universal property; but when it can be used, it often gives rings in which one can compute more easily than in rings arising by universal localization.

P. M. Cohn has discovered that many of the deficiencies of universal localization mentioned earlier can be overcome by inverting, not *elements* of R, but *matrices*. Cf. his books mentioned at the end of the comment on specializations, and for another approach to the material, my paper *Constructing division rings as module-theoretic direct limits,* Trans. A. M. S. **354** (2002) 2079-2114.

The fact that we are studying fields of fractions will be my excuse for mentioning a few simple but long-outstanding questions about rational numbers:

Open Question II.4.c2 (Richard Guy)**.** *Does there exist a point of the plane having rational distances from each of the vertices of the unit square,* (0,0), (0,1), (1,0), (1,1)? (= Problem D19 in Richard K. Guy, *Unsolved Problems in Number Theory,* Springer-Verlag, 1981.)

Open Questions II.4.c3 (Erdős and Graham)**.** (a) *For every positive integer* $d > 1$, *do there exist positive integers* d_1, d_2, d_3 *such that*

$$4/d = 1/d_1 + 1/d_2 + 1/d_3 ?$$

A more sweeping question is

(b) *For every positive integer* n, *is it true that all but* finitely many *of the rational numbers* n/d $(d > 0)$ *can be written* $1/d_1 + 1/d_2 + 1/d_3$ *as above?* (Cf. p.44 of P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory,* Université de Genéve, 1980, and references given there.)

Re §II.5. Principal and factorial rings.

P.111, definition of **factorial ring** [~]: The standard term in English, which we shall use, is *unique factorization domain*, abbreviated UFD. (Lang almost gives this in his parenthetical alternative ''unique factorization ring''.) Recall also that we will be using the standard terms *integral domain* and *principal ideal domain* (PID) where Lang says ''entire ring'' and ''principal entire ring''.

P.112, second line [=]: By (a, b) Lang means $(a) + (b)$. I dislike the (a, b) notation, since it looks just like the symbol for an ordered pair; but it is common among number theorists and others. However, paging around this book, I don't see any other occurrences of it before Chapter IX.

P.112, fifth line after second display [=]: The conclusion ''and has a maximal element'' is by Zorn's Lemma.

P.112, proof of Theorem II.5.2 [~]:

The Ascending Chain Condition, and Theorem II.5.2.

The proof of Theorem II.5.2, that every PID is a UFD, becomes more understandable with some background. Let us note first that the hypothesis that A is a PID can be broken into two parts, which are used in very different ways. On the one hand it tells us that every ideal generated by *two* elements is principal, from which, by induction, we see that every ideal generated by finitely many elements is principal. Given this observation, the remainder of the condition that *every* ideal is principal is clearly just the statement that every ideal is finitely generated.

Now the condition that every 2-generator ideal is principal tells us that any two elements not only have a g.c.d., but a g.c.d. which can be written as a linear combination of these elements. In particular, if p is an irreducible element and a an element not divisible by p, then irreducibility of p says that p and a

have no nonunit common divisor, so their g.c.d. is 1, and we can write this as a linear combination $1 = xp + ya$. Taking the image in $A/(p)$ of this equation, we see that a becomes invertible in that ring. Thus, *every* nonzero element of $A/(p)$ is invertible, i.e., $A/(p)$ is a field. This means that (p) is a *maximal* ideal, and in particular, is *prime*; i.e., that if a product of two elements is divisible by p, one of these elements is divisible by p. This is the only consequence of the condition ''every 2-generator ideal is principal'' that we use.

The other half of the PID condition constitutes an extremely important concept of ring theory, which we characterize in the following

Lemma II.5.c1. *The following conditions on a commutative ring A are equivalent:*

(i) *Every ideal of A is finitely generated.*

(ii) *A has no infinite strictly ascending chain of ideals $\mathbf{a}_1 \subsetneqq \mathbf{a}_2 \subsetneqq \dots$.*

(iii) *Every nonempty set of ideals of A has a maximal element.*

Proof. (i) $\Rightarrow$ (ii): Assuming we have an infinite chain $\mathbf{a}_1 \subsetneqq \mathbf{a}_2 \subsetneqq \dots$, choose a finite generating set $x_1, \dots, x_n$ for its union. Each x_i lies in some $\mathbf{a}_{j_i}$; let $j = \max(j_1, \dots, j_n)$. Then clearly $\mathbf{a}_j = \mathbf{a}_{j+1}$, a contradiction.

(ii) $\Rightarrow$ (iii): Let S be a nonempty set of ideals of A. Take $\mathbf{a}_1 \in S$; if it is not maximal, choose $\mathbf{a}_2$ strictly larger; if this is not maximal, choose $\mathbf{a}_3$ strictly larger than $\mathbf{a}_2$, and so on. If this process did not terminate in finitely many steps, we would get a contradiction to (ii), so it terminates, i.e., at some step we get a maximal member of S.

(iii) $\Rightarrow$ (i): Given an ideal $\mathbf{a} \subseteq A$, let $\mathbf{b}$ be a maximal element in the set of finitely generated subideals of $\mathbf{a}$. We claim $\mathbf{b} = \mathbf{a}$. For if not, we could take $x \in \mathbf{a} - \mathbf{b}$, and then $\mathbf{b} + xA$ would be a larger finitely generated subideal of $\mathbf{a}$, contradicting our choice of $\mathbf{b}$. □

A ring satisfying the equivalent conditions of the above lemma is called *Noetherian* (after Emmy Noether. Cf. Lang, introductory paragraph of §IV.4.) Condition (ii) is expressed by saying that A has *ascending chain condition*, abbreviated ACC; or more precisely, ACC on ideals. Clearly, the condition of ACC can be defined for arbitrary partially ordered sets. The analog of the lemma holds for essentially any sort of algebraic structure, with ''ideal'' replaced by any class of substructures which is closed under taking unions of chains.

Just as the fact that the natural numbers are well-ordered, i.e., satisfy *descending chain condition*, gives us the principle of mathematical induction, so one can do induction on any partially ordered set with ascending chain condition. Let me formulate this for Noetherian rings as

Lemma II.5.c2 (Principle of Noetherian induction)**.** *Let A be a Noetherian ring and P a condition on ideals, such that if $\mathbf{a}$ is an ideal and all ideals strictly larger than $\mathbf{a}$ satisfy P, then $\mathbf{a}$ satisfies P. Then all ideals of A satisfy P.*

Proof. Assuming the contrary, let S be the set of ideals of A that do not satisfy P. By the Noetherian assumption, S has a maximal element $\mathbf{a}$. Then all ideals strictly larger than $\mathbf{a}$ satisfy P, hence, by assumption, $\mathbf{a}$ too must satisfy P, a contradiction. □

The effect of this principle is that in trying to prove a result about ideals in a Noetherian ring, one can safely say ''Let $\mathbf{a}$ be an ideal, and assume by Noetherian induction that all larger ideals satisfy the desired property.'' If one can deduce from this that $\mathbf{a}$ also has the desired property, then one has established that all ideals do.

The same method can be applied to any subclass of the ideals of a Noetherian ring. E.g., if we want to prove that all *principal* ideals have a property P, we may apply Noetherian induction to the condition

''either $\mathbf{a}$ is nonprincipal, or $\mathbf{a}$ satisfies P''.

Note that any PID satisfies condition (i) of Lemma II.5.c2, and so is Noetherian. Returning now to Lang's proof that every PID is a UFD, we can replace the explicit argument concerning an ascending chain of principal ideals by the reasoning, ''To prove that every nonzero nonunit $a \in A$ has a factorization into irreducibles, assume by Noetherian induction, applied to the nonzero principal ideals, that every nonunit that generates a larger ideal than (a) can be so factored. Now if a is itself irreducible, there is nothing to prove, while if there is a nontrivial factorization $a = bc$, then b and c each generate larger ideals than a, so by our inductive hypothesis, each may be factored into irreducibles. Multiplying together these factorizations, we get a factorization of a into irreducibles, as required.''

Knowing that every nonzero nonunit factors into irreducibles, and that every irreducible is prime, the proof of unique factorization is now easily completed as on the bottom of p.112.

For the reader who has, as recommended in the comments on §II.3, read the beginning of the chapter on modules, I will mention that an ideal of a commutative ring R (or a left ideal of any ring R) is the same thing as a (left) R-submodule of R, making the concept of a Noetherian ring a special case of that of a Noetherian module. Lemma II.5.c1 above is stated and proved in this broader context on the first two pages of Chapter X. (If you are interested, you should find all of §X.1 readable at this point.)

P.113, second and fourth paragraphs (''We could call ...'' and ''We observe ...'') [<]: Here Lang is talking about UFD's (''factorial rings''); though he doesn't state that assumption, he uses it.

P.113, line 3 [=]: ''We could call two elements $a, b \in A$ equivalent if there exists a unit u such that $a = bu$''. In fact, such elements a and b are standardly called *associates*. (This definition applies in any commutative ring, not necessarily a UFD.) If A is an integral domain, then elements x and y are associates if and only if $xA = yA$. (If A is not an integral domain, this is no longer true. The problem of finding a counterexample is Exercise II.5:4.)

P.113, definition of **least common multiple** [~]: Though Lang has been talking about unique factorization domains, one can give a definition of ''least common multiple'' in an arbitrary integral domain, analogous that of greatest common divisor (p.111 bottom); namely, a least common multiple of a finite family of elements means an element m which is a multiple of all the members of the family, and such that every common multiple of the family is a multiple of m. In a unique factorization domain, one sees that the element Lang describes here has this property. In a general integral domain, a finite family of elements may or may not have a least common multiple or a greatest common divisor.

P.114, end of §II.5 [>].

Sums and intersections of ideals, and g.c.d.'s and l.c.m.'s of elements.

If $\mathbf{a}$ and $\mathbf{b}$ are two ideals of a commutative ring A, we see that their sum $\mathbf{a}+\mathbf{b}$ is the least ideal containing both. There is also a greatest ideal contained in both, namely $\mathbf{a} \cap \mathbf{b}$. Note that the product ideal, $\mathbf{ab}$, is clearly contained in this intersection; it may or may not equal it. In a principal ideal domain, it is not hard to show that the product is strictly smaller than the intersection whenever $\mathbf{a}$ and $\mathbf{b}$ are generated by elements having a nontrivial common divisor. On the other hand, in *any* commutative ring A, I claim that if $\mathbf{a}$ and $\mathbf{b}$ are ideals such that $\mathbf{a} + \mathbf{b} = A$, then $\mathbf{a} \cap \mathbf{b} = \mathbf{ab}$. We have seen ''$\supseteq$''; to get ''$\subseteq$'', note that for any two ideals of a commutative ring, we have

$$(\mathbf{a} \cap \mathbf{b})(\mathbf{a}+\mathbf{b}) \;=\; (\mathbf{a} \cap \mathbf{b})\mathbf{a} + (\mathbf{a} \cap \mathbf{b})\mathbf{b} \;\subseteq\; \mathbf{ba} + \mathbf{ab} \;=\; \mathbf{ab}.$$

Substituting in $\mathbf{a} + \mathbf{b} = A$ gives the desired result. (This was stated as an ''exercise for the reader'' on p.87, third line from the bottom.)

Observe that a ''greatest common divisor'' of elements a and b of a commutative ring A means, in effect, an element c such that the ideal cA is the *smallest* principal ideal containing aA and bA; equivalently, the smallest principal ideal containing $aA + bA$. (As we have noted, given a and b, such a c may or may not exist.) In particular, if $aA + bA$ is principal (e.g., if A is a principal ideal

domain), then a g.c.d. of a and b means a generator of this ideal. However, a g.c.d. may exist even if the ideal $aA + bA$ is not principal. In the polynomial ring $\mathbf{R}[x, y]$, we find that 1 is a g.c.d. of the elements x and y, though the ideal they generate consists of all polynomials with zero constant term, and so does not contain 1.

A ''least common multiple'' of a and b similarly means an element c such that cA is the greatest principal ideal contained in $aA \cap bA$. Since every ideal is a union of the principal ideals it contains, in this case we see that the only way such a greatest principal ideal can exist is if $aA \cap bA$ is itself principal.

Chapter III. Modules.

Re §III.1. Basic definitions.

P.117, bottom [>]:

The idea of a module.

The relationship between rings R and R-modules M is analogous to that between groups G and G-sets S which we studied in §I.5. Just as the concept of a group is motivated by that of a group of permutations of a set S, and the abstract concept of a group is brought back to its roots by studying sets S on which a group G acts, equivalently, sets S given with group homomorphisms $G \to \mathrm{Perm}(S)$, so the concept of a *ring* is motivated by the properties of the set of *endomorphisms of an abelian group*, and an important role in ring theory is played by abelian groups M on which a ring R acts by endomorphisms; equivalently, abelian groups M given with ring homomorphisms $R \to \mathrm{End}(M)$. These are called *left R-modules*.

In the study of actions of groups, though we set up the definition so that elements of the group were written to the left of elements of the set, there are some cases where it is more natural to write the group elements on the right. However, there is no essential difference between ''right'' and ''left'' actions, because if we have an action of one sort, we can re-encode it as an action of the opposite sort, using the observation that the group operation $(\ )^{-1}$ reverses the order of multiplication. (In formal terms $(\ )^{-1}$ is an ''anti-automorphism'' of any group G.) Thus, given a ''right action'', under which a group element g takes an element x to an element $x \cdot g$, satisfying $((x)\cdot g)\cdot h = x\cdot(gh)$, though we cannot define a left action by writing $gx =_{\text{def.}} x\cdot g$, because this definition will not behave properly with respect to composition, the definition $gx =_{\text{def.}} x\cdot g^{-1}$ will. In ring theory, on the other hand, a ring R may not have an anti-automorphism, so the concepts of *left R-module* and *right R-module* are not equivalent. However, right modules over a ring R are equivalent to left modules over a ring called R^{op} (''the opposite ring to R'') having the same elements and additive structure as R, but the reverse multiplication, $a * b = ba$. Thus it will suffice to discuss left modules almost exclusively in presenting general results. When we deal with modules over commutative rings, there is no distinction between left and right modules.

Since every abelian group has endomorphisms $x \mapsto nx$ for every integer n, every abelian group has a natural structure of $\mathbf{Z}$-module; indeed, $\mathbf{Z}$-modules are essentially the same thing as abelian groups. So you can think of the theory of modules as a generalization of the theory of abelian groups. Another class of modules with which you are already familiar are modules over a field, which are called *vector spaces*.

Incidentally, in defining a module, Lang referred to ''an operation of A on M (viewing A as a multiplicative monoid ...)''. Actually, he hasn't defined the concept of an action of a monoid on a set, but the definition is analogous to that of the action of a group on a set. The latter was defined as a group homomorphism $G \to \mathrm{Perm}(S)$; so the former may be defined as a monoid homomorphism $G \to \mathrm{Map}(S)$. Either concept is also equivalent to a map $G \times S \to S$ with the properties $1s = s$ and $(gh)s = g(hs)$ for

all $g, h \in G$, $s \in M$.

P.118, end of middle block of paragraphs [>]:

Some basic constructions. Let M be a left R-module. Then you should check that (i) for any subset $N \subseteq M$, the set of elements of R which annihilate N (i.e., $\{r \in R \mid rN = 0\}$) is a left ideal of R; (ii) for any additive subgroup $N \subseteq M$, the set of elements of R carrying M into N will form a *right* ideal, and (iii) under the same assumption, the set of elements which carries N into *itself* will be a *subring*. Moreover, (iv) if N is in fact a submodule of M, then its annihilator (which we noted in (i) above was a left ideal), and the set of elements of R carrying M into N (which we noted in (ii) was a left ideal) will both be 2-sided ideals of R.

Note also in connection with (i) that if N is a singleton $\{x\}$, then the structure of the submodule $Rx = \{rx \mid r \in R\} \subseteq M$ is determined by the annihilator ideal I of x: $Rx \cong R/I$; just as the structure of the orbit Gx of an element x of a G-set is determined by the isotropy subgroup of x: $Gx \cong G/G_x$.

The above four constructions are useful ''generic'' ways of getting one-sided and two-sided ideals and subrings of R. Indeed, it is instructive to verify that *every* right ideal, left ideal, 2-sided ideal, and subring of R can be obtained in the ways indicated above, using appropriate M and N. And, just as it can be heuristically useful to regard a general subgroup H of a group G as the isotropy subgroup of an element of a G-set, so the above constructions give useful ways of thinking about ideals and subrings of a ring.

P.118, paragraph containing the bold-face words **vector space** [>]: Whenever we construct a ring R as a ring of endomorphisms of an abelian group or module M, that very construction makes M an R-module (what W.-Y. Hsiang likes to call the ''birth certificate representation'' of the given ring). Thus, since the ring of $n \times n$ matrices over K, often denoted $M_n(K)$, is constructed to model the ring of linear endomorphisms of the vector space K^n, that vector space has, as Lang notes, a structure of $M_n(K)$-module.

Likewise, since the Weyl algebra (described in my comments on §II.1) arises as the algebra of operators on a polynomial algebra $\mathbf{R}[x]$ generated by the operators X of multiplication by x, and Y of differentiation with respect of x, it follows that one can make $\mathbf{R}[x]$ into a module over this algebra by letting X and Y act in these ways.

Since I recommended reading sections III.1-III.4 before §II.3, we don't have many examples of rings yet, making it hard to give more examples of modules! At the end of §II.3 I note some further examples of modules.

P.119, last two lines, and p.120, top line [~]: If $f : M \to M'$ is a module homomorphism, then the module $M'/f(M)$ is universal (initial) among modules given with a homomorphism from M' into them whose composite with f is the zero homomorphism. This property is dual to that of the *kernel* of f, which is universal (terminal) among modules with a homomorphism into M whose composite with f is zero; hence $M'/f(M)$ is called the *cokernel* of f. We have the usual terminological alternatives of speaking loosely of the module $M'/f(M)$ as the cokernel, or using this term more precisely for the universal pair consisting of this module and the canonical map of M' into it. Lang mentions both usages, except that, as usual, he would abbreviate the pair to the map alone.

P.120, bottom, definitions of **monomorphism** and **epimorphism** [>]:

The question of what the words *monomorphism* and *epimorphism* should mean is currently a can of worms! They were originally coined (I believe by Bourbaki) as terms for ''one-one homomorphism'' and ''surjective homomorphism'' respectively. However, the early category-theorists, looking for ''category-theoretic equivalents'' of various concepts, noted that in almost all naturally occurring categories, the one-to-one homomorphisms could be characterized as those homomorphisms $f : A \to B$ such that for every object X and every pair of homomorphisms $i, j : X \to A$, one has $fi = fj \Rightarrow i = j$. They therefore named a morphism f in a general category having this property a ''monomorphism''. They then felt it natural to define an ''epimorphism'' to mean a morphism $f : A \to B$ with the dual property, namely that

for every object Y and pair of morphisms $i, j: B \to Y$, one has $if = jf \implies i = j$.

Now it turns out that in about half the naturally occurring categories, the class of epimorphisms, so defined, coincides with the class of surjective homomorphisms, while in the other half it does not! However, in the latter cases, the class of epimorphisms *does* constitute a class of homomorphisms worth studying, and there has been considerable work on characterizing such maps among various sorts of algebras. Hence, ''epimorphism'' is now frequently used by algebraists in its category-theoretic sense, and I do *not* recommended using the term in the old meaning. The older meaning can conveniently be expressed by the phrase ''surjective homomorphism'' or ''onto homomorphism'', and I will do so here.

The category of modules over a ring R happens to be one of those in which epimorphisms are the same as surjective homomorphisms, so Lang's use of the term on p.120 does not contradict the category-theoretic meaning. However, I will generally use the more down-to-earth terms when there is a choice, unless the context is category-theoretic.

How should the statement that a morphism $f : X \to Y$ of algebraic objects is an epimorphism be thought of, in cases where it is not equivalent to surjectivity? Intuitively, it means that the subobject $f(X)$ ''controls'' all of Y. Depending on the sort of object we are dealing with, this may or may not entail $f(X) = Y$, i.e., surjectivity. (See Exercise I.11:1 for some positive and negative examples.) On the other hand, surjective homomorphisms are always epimorphisms. (Note that though the concepts of monomorphism and epimorphism are defined in arbitrary categories, these remarks on their relation with one-one-ness and onto-ness refer to categories of algebraic objects, and homomorphisms that are maps of underlying sets. As we have observed, one-one-ness and onto-ness are not even defined for objects of general categories.)

Although it is rare for *monomorphisms* to be anything other than one-to-one maps, here is one example where this, too, fails. Let $\mathbf{C}$ be the category of *divisible* abelian groups, that is, abelian groups which for every positive integer n satisfy $(\forall\, x)(\exists\, y)\ ny = x$. You should not find it hard to prove that in this category, the canonical map $\mathbf{Q} \to \mathbf{Q}/\mathbf{Z}$ is a monomorphism.

P.121, first paragraph [~]:

Notes on the concept of an algebra.

Lang's introduction to this concept is misleading; let me put it in context. There are many ways one can generalize the concept of *ring*. On the one hand, one can remove the associativity assumption on the multiplication, or the assumption of the existence of unit. The resulting objects are called *nonassociative rings* and *nonunital rings*. (In the former case, one frequently restricts attention to rings which, instead of associativity, satisfy some other interesting identity.)

On the other hand, one may note that in any ring, the underlying additive group structure can be looked at as a structure of $\mathbf{Z}$-module, but that in some naturally occurring examples, it has a natural structure of module over a larger ring, e.g., the rationals, the reals, or the complex numbers. One calls a ring given with a structure of A-module on its additive group, satisfying appropriate conditions, an *A-algebra*. What are these appropriate conditions? Note that in a ring R one can prove (from distributivity) that for any ring elements x and y and any integer n, one has $(nx)y = n(xy) = x(ny)$. To make the concept of A-algebra a good generalization of that of ring, one requires, likewise, that for any elements x and y of the algebra, and element $a \in A$, one have

(c19) $$(ax)y \;=\; a(xy) \;=\; x(ay).$$

(This is part of the condition in Lang, that the multiplication map of E be *bilinear*.) Note that if $a, b \in A$, $x, y \in R$, then (c19) can be used to transform $(ax)(by)$ either to $(ab)(xy)$ or to $(ba)(xy)$. Thus these must be equal. This leads to a degenerate theory unless A is assumed commutative; hence the concept of A-algebra is only defined for commutative rings A.

The two sorts of generalization of the concept of ring that we have discussed – dropping unitality or

associativity on the one hand, and enhancing the underlying abelian group structure to an A-module structure on the other – are independent of one another. Thus, there is no good reason why Lang has connected the transition from rings to A-algebras with the possible removal of the conditions of associativity and/or unitality, as though rings were required to have these properties, but algebras were not.

Let us make explicit an observation that is particular to the case of associative unital algebras, and gives another way of looking at these animals. (We will use the words ''associative unital'' here, because we have been talking about other sorts of rings and algebras; but elsewhere, we will understand ''ring'' and ''algebra'' to include these conditions unless the contrary is stated.)

Lemma III.1.c1. *Let A be a commutative ring. Then the following structures are equivalent.*

(a) *An associative unital A-algebra R, i.e., an A-module R given with an associative unital bilinear multiplication.*

(b) *An associative unital ring R, together with a homomorphism f from A into the center of R.*

Sketch of Proof. Given an associative unital A-algebra R, if we regard its multiplication map as a bilinear map of abelian groups, we see that it makes R into a ring. Now let $f: A \to R$ be the map taking $a \in A$ to $a1 \in R$ (where 1 denotes the unit of the ring R, and this is multiplied by $a \in A$ using the A-module structure of R). It is easy to verify that this gives a homomorphism from A into the center of R.

Inversely, given an associative unital ring R and a homomorphism f from A into the center of R, we may make R an A-module by defining the product of $a \in A$ and $r \in R$ to mean the product $f(a)r$, evaluated in the ring R. It is easy to verify that this makes R an A-module, such that the ring multiplication of R is A-bilinear. Thus, R becomes an A-algebra, and the conditions of associativity and unitality hold in this algebra because they hold in R as a ring.

Finally, it is easy to verify that if we pass from a structure of unital associative A-algebra to a structure of ring given with a map of A into its center, and then back to an algebra structure, in the manner indicated, we get precisely the structure we started with; and likewise if we go from a structure of ring with a map of A into its center to an algebra structure and back. Thus, the constructions we have defined are inverse to one another, and give a bijection between the two sorts of structure on a set R. □

(Lang sketches the above idea still more briefly in the next-to-last paragraph of p.121.)

Note that if R is commutative, the *center* of R, referred to in (b) above, is all of R. Hence we have

Corollary III.1.c2. *Let A be a commutative ring. Then the following structures are equivalent.*

(a) *A commutative associative unital A-algebra R.*

(b) *A commutative associative unital ring R, together with a homomorphism $f: A \to R$.* □

There is another natural motivation for the concept of a unital associative A-algebra. If A is a commutative ring, and M is an A-module M, what structure does the set $\mathrm{End}_A(M)$ of all A-module endomorphisms of M have? Its elements can be added; they can be composed with each other, *and they can be multiplied by elements of A*. The first two of these operations satisfy the axioms for an associative unital ring, and the multiplication of this ring is bilinear with respect to the action of A. So we may think of the concept of A-algebra as an abstraction of the natural structure on these sets of A-module endomorphisms, just as the concept of ring abstracts the natural structure on the set of endomorphisms of an abelian group.

There is, unfortunately, another important technical use of the word ''algebra''. The subject-matter of the wide branch of mathematics called ''algebra'' is the properties of sets given with one or more operations – e.g., monoids, groups, G-sets for a fixed group G, rings, modules, etc.. Though many approaches to algebra, including that of this course, treat these as separate topics and introduce them one

by one, the field called *Universal Algebra* or *General Algebra* studies the properties that such concepts have in common, and calls *any* such object an *algebra*. It would be very desirable if there were a name for this concept, distinct from the term used for the specific concept studied in ring-theory, but no one has come up with a good one, and the double usage is now entrenched. When there is ambiguity, one must distinguish the two concepts as ''an algebra in the sense of universal algebra'' and ''an algebra in the sense of ring theory''. Russian authors often call the former ''a universal algebra'', but this is not a good choice, because no universal property is assumed. In this course, unless otherwise stated, the word ''algebra'' will be used only in its ring-theoretic sense.

Lang does not say what is meant by a homomorphism of A-algebras. If we regard an A-algebra as an A-module with a ring structure whose multiplication is A-bilinear, then a homomorphism of A-algebras means a map which is both a ring homomorphism and an A-module homomorphism. If we regard an A-algebra as a ring with a homomorphism of A into its center, then a homomorphism of A-algebras means a ring homomorphism which makes a commuting triangle with the homomorphisms from A. These two definitions are easily shown equivalent.

What about modules over algebras? There are two ways of formulating the concept, but fortunately, they come to the same thing.

Lemma III.1.c3. *Let A be a commutative ring, and R an A-algebra. Then the following structures are equivalent:*

(a) *A left R-module M, that is, an abelian group given with a ring homomorphism $R \to \mathrm{End}(M)$, (where $\mathrm{End}(M)$ denotes the ring of abelian group endomorphisms of M).*

(b) *An A-module M given with an A-algebra homomorphism $R \to \mathrm{End}_A(M)$, (where $\mathrm{End}_A(M)$ denotes the A-algebra of A-module endomorphisms of M).*

Sketch of Proof. It is clear that a structure as in (b) gives, in particular, a structure as in (a), since $\mathrm{End}_A(M)$ is a subring of $\mathrm{End}(M)$, and an algebra homomorphism into the former is, in particular, a ring homomorphism into the latter. On the other hand, given a structure as in (a), note that the map of A into the center of R, composed with the map $R \to \mathrm{End}(M)$ gives a map $A \to \mathrm{End}(M)$, i.e., an A-module structure on A. The action of each $r \in R$ commutes with the actions of all elements of A (why?), hence each such element in fact gives an A-module endomorphism of M; so we get a structure as in (b). □

Recall that a $\mathbf{Z}$-module is the same as an abelian group. From this we easily see that a $\mathbf{Z}$-algebra is the same as a ring. Hence if we prove a result about algebras, we also have the corresponding result for rings.

P.121, **Examples** [<]: Skip this example, unless you are familiar with the subject matter. (I would love to insert here a thumbnail sketch of the concept of Lie algebra, but I think I've loaded in enough commentary on this page!)

Students who are in my Math 250A or others who are following my suggestion of reading this material immediately after §II.2 should also skip the discussion of group and monoid algebras; these will be introduced in §II.3.

P.121, next-to-last paragraph [~]: This is stated more precisely in Lemma III.1.c1 above.

Re §III.2. The group of homomorphisms.

P.122, second paragraph [>]: You should convince yourself that the calculation showing that $\mathrm{Hom}_A(X', X)$ becomes an A-module indeed works if A is commutative, but not for noncommutative A. (You might find it interesting to think about the question: Given a noncommutative ring A, is there in general *some* structure stronger than that of an abelian group that we can put in a functorial way on these hom-sets?)

Caveat: The sum of two ring homomorphisms is *not* in general a ring homomorphism. For rings R

and S, the set $\mathrm{Hom}(R, S)$ has no algebraic structure.

P.122, Proposition III.2.1 and p.123, Proposition III.2.2 [~]: Short exact sequences are a way of indicating the relation between a module, a homomorphic image of this module, and the kernel of the homomorphism. Since these are key module-theoretic constructions, when we study functors on module-categories, one of the first things we would like to know is how they behave on such sequences. We see from these two propositions that functors of the form $\mathrm{Hom}(X, -)$ and $\mathrm{Hom}(-, Y)$ each preserve most, but not all, of the exactness of short exact sequences. You should, in fact, look for examples showing that if

$$0 \to M' \to M \to M'' \to 0$$

is a short exact sequence of modules (e.g., $\mathbf{Z}$-modules, i.e., abelian groups), neither of the sequences

$$0 \to \mathrm{Hom}(X, M') \to \mathrm{Hom}(X, M) \to \mathrm{Hom}(X, M'') \to 0$$

$$0 \leftarrow \mathrm{Hom}(M', Y) \leftarrow \mathrm{Hom}(M, Y) \leftarrow \mathrm{Hom}(M'', Y) \leftarrow 0$$

need be exact, at the place where the respective proposition does not guarantee exactness.

The statements of the propositions are not limited to the cases where the given sequence is part of a short exact sequence; i.e., where the leftmost map of Proposition III.2.1 is one-to-one, and the rightmost map of Proposition III.2.2 is onto. However, you might verify that if a covariant or contravariant functor from modules to abelian groups preserves exactness of all sequences obtained by dropping the appropriate zero-module from a genuine short exact sequence, then it preserves exactness for the more general sort of exact sequence referred to in these propositions. Another useful result you might verify is that any functor from A-modules to abelian groups which takes all short exact sequences to short exact sequences in fact takes all exact sequences to exact sequences.

The fact that $\mathrm{Hom}(X, -)$ preserves exactness of sequences $0 \to Y' \to Y \to Y''$ is expressed by saying that it is a *left exact functor*; the fact that $\mathrm{Hom}(-, Y)$ preserves exactness of sequences $X' \to X \to X'' \to 0$ (though reversing the direction of arrows) is expressed by saying that it is a *right exact* (*contravariant*) *functor*.

In his proofs of these two results, Lang leaves to the reader, among other things, the ''upward'' direction; that is, the verification that *if* a sequence of the sort shown has the property that the induced sequences of hom-groups are exact for all choices of the outside object, *then* the given sequence is itself exact. The needed arguments are easy to find – once one is familiar with the following heuristic principle:

(c20) *To study the consequences of a condition holding for all morphisms of a given sort, consider a universal example.*

Let us see how to apply this in proving the ''upward'' direction in Proposition III.2.1. Lang doesn't name the morphism $X'' \to X$ of the given sequence; let us call it μ. One of the conditions for exactness of the bottom sequence of hom-sets is that the kernel of the left-hand arrow be equal to the image of the right-hand arrow. To say that it *contains* this image for every Y is to say that for every Y, every element of the right-hand hom-set, $\mathrm{Hom}_A(X'', Y)$, when composed with μ (to map it into the middle hom-set) and then with λ (to map it into the left-hand hom-set) gives 0. So let us apply this observation to a *universal* example of a morphism from X'' into a module Y. Given no other conditions on the morphism than that it have domain X'', a universal example is the identity morphism of X''. The statement that this identity morphism, when composed with μ and then λ, gives 0 is just the condition $\mu\lambda = 0$, equivalently, $\mathrm{Im}(\lambda) \subseteq \mathrm{Ker}(\mu)$, which is indeed part of the desired statement that the original sequence be exact.

Let us now treat similarly the condition that for all Y, the kernel of the left-hand arrow *be contained in* the image of the right-hand arrow. This is a statement about every element of that kernel, so we should look for a universal instance of an element of that kernel; i.e., a universal example of a homomorphism

from X into a module Y which goes to zero on composition with the map $\lambda: X' \to X$. If you draw a picture, you will see that the universal map with this property is the canonical map

$$q: X \to \mathrm{Coker}(\lambda).$$

So let us apply the exactness assumption on the bottom sequence in the case $Y = \mathrm{Coker}(\lambda)$. Since q is then in the kernel of the left-hand map of that sequence, there must be an element p in the right-hand hom-set $\mathrm{Hom}(X'', \mathrm{Coker}(\lambda))$ that maps to q in the middle hom-set, i.e., such that $q = p\mu$. Hence $\mathrm{Ker}(\mu) \subseteq \mathrm{Ker}(q) = \mathrm{Im}(\lambda)$ (the latter equality by the definition of q). Since we proved $\mathrm{Im}(\lambda) \subseteq \mathrm{Ker}(\mu)$ in the preceding paragraph, we get $\mathrm{Im}(\lambda) = \mathrm{Ker}(\mu)$, as desired.

The other parts of the ''upward'' direction of this proposition, and the proof of the same direction in the next proposition, are similar applications of (c20), which you should work out to get familiarity with this technique.

P.124, next-to-last paragraph [>]: There is no need to assume A commutative in this paragraph.

Note that, given what Lang calls a representation of R on an A-module M, the full structure one has on M can be described as one of an abelian group, together with a homomorphism $\alpha: A \to \mathrm{End}(M)$ giving the A-module structure, and a homomorphism $\rho: R \to \mathrm{End}(M)$ giving the action of R, subject to the condition that for all $a \in A$, $r \in R$, $x \in M$ we have $\alpha(a)\,\rho(r)(x) = \rho(r)\,\alpha(a)(x)$; or, writing the actions both of elements of R and of elements of A as ''multiplications'', $arx = rax$.

However, noncommutative ring theorists frequently make a slight notational change, writing endomorphisms of a *left* A-module M on the *right* of M, e.g., as $x \mapsto xh$, and defining composition of endomorphisms accordingly, so that $x(hh') = (xh)h'$. Let us denote the ring of such endomorphisms by $\mathrm{End}({}_AM)$; this will be the *opposite* of the ring Lang denotes $\mathrm{End}_A(M)$. (The *opposite ring* is defined in the comment to p.117 of Lang in this Companion.) Thus, given what Lang calls a representation of the ring R on the A-module M, if we write B for the opposite of the ring R, we can describe the map giving the representation as equivalent to a ring homomorphism $B \to \mathrm{End}({}_AM)$; in other words, a way of letting elements of B act on the *right* on the A-module M. The condition that the actions of elements of B respect the actions of elements of A now takes the form

$$(ax)b = a(xb) \qquad (a \in A,\ r \in R,\ x \in M).$$

An abelian group with actions of A and B on the left and the right respectively, and satisfying this law, is called an (A,B)-*bimodule*. (The name refers to the fact that it is composed of a *left A-module* structure and a *right B-module* structure, related by the above condition.) A major advantage of this notation is that if A and B have elements in common (for instance, if they are the same ring), then to unambiguously write the actions of both on the left would require awkward notation like $\alpha(a)\beta(b)x$; while the notation axb is unambiguous, and, in view of the above displayed equation, does not even require parentheses.

An important class of bimodules is given by the 2-sided ideals of a ring A; these are clearly (A,A)-bimodules in a natural way. (Note that Lang's notation would require calling these ''representations of A^{op} on left A-modules''.)

You will sometimes see the term (A,B)-bimodule written ''$A-B-$bimodule'', but I do not recommend this, since if in place of A and B one has complicated symbols, it can be unclear how the prefix is divided. We will not study bimodules (or ''representations'') in 250A, but you should be aware of the concept.

I have some further discussion of the convention of writing endomorphisms of left modules on the right and vice versa – its advantages and disadvantages – in a note at
http://math.berkeley.edu/~gbergman/grad.hndts/left+right.ps .

Re §III.3. Direct products and sums of modules.

P.129, 7 lines from bottom [~]: Lang says an A-module generated by a subset S may be written $A\langle S\rangle$. Don't! As I have mentioned, that is noncommutative ring-theorists' notation for the *ring* generated over A by a set of elements S of an overring. For the A-module generated by S, simply write AS.

P.129, long paragraph, containing definition of submodule **generated** by a set [>]: This construction thus takes its place beside ''the submonoid of a monoid G generated by a subset $X \subseteq G$'', ''the subgroup of a group G generated by a subset $X \subseteq G$'', ''the normal subgroup of a group G generated by a subset $X \subseteq G$'', ''the subring of a ring R generated by a subset $X \subseteq R$'' and ''the right (or left, or 2-sided) ideal of a ring R generated by a subset $X \subseteq R$''. Note that in each of these cases, the set constructed is obtained as the *closure* of X under certain operations. Some properties common to such constructions are abstracted by the concept of

Closure operators.

Suppose that in each of the cases mentioned in the preceding paragraph, we denote the substructure ''generated by X'' as $\mathrm{cl}(X)$. Clearly, in each case we have

$$\mathrm{cl}(X) \supseteq X, \tag{c21}$$

$$X \subseteq Y \Rightarrow \mathrm{cl}(X) \subseteq \mathrm{cl}(Y), \tag{c22}$$

$$\mathrm{cl}(\mathrm{cl}(X)) = \mathrm{cl}(X). \tag{c23}$$

Operations cl with these properties are ubiquitous, not only in algebra, but in other areas of mathematics. (For instance, if S is any topological space, the operation of taking the *topological closures* of subsets of S satisfies these conditions, while if S is Euclidean n-space, the operation taking a set to its *convex hull* has the same properties.) An operator cl on subsets of any set S that satisfies (c21)-(c23) is called a *closure operator*, and a subset X of S is called *closed* with respect to cl if $\mathrm{cl}(X) = X$.

It is easy to verify that if cl is a closure operator on a set S, then the *intersection* of any family of closed subsets of S is a closed subset, and the improper subset $S \subseteq S$ is closed. Conversely, if C is a family of subsets of a set S such that the intersection of any set of elements in C is in C, and S itself is in C, then there is a unique closure operator cl on S whose closed sets are the members of C; namely, the operator taking every $X \subseteq S$ to the intersection of all members of C containing X. So closure operators on S are equivalent to sets of subsets with the above properties.

Particular sorts of closure operators satisfy additional conditions. For instance, examples arising in algebra, including all those that have come up in this course, typically satisfy the condition that the union of a *chain* of closed subsets is closed, while this is not true of topological closure. On the other hand, topological closure has the property that the union of finitely many closed sets is closed, which the algebraic examples do not.

At this point, we will not study the general concept any further, but simply put this useful definition under our belts. We will see it again in our discussions of §VI.1, §VI.14 and §XI.1.

P.129, next-to-last paragraph [~]: Lang indicates that a finitely generated module can also be called a ''finite'' module. This terminology is used by some algebraic geometers, but since it can lead to confusion, I don't recommend it, unless you are talking with those who use it regularly.

P.130, middle [>]: The concepts of linear independence for *sets* and *families* are both important, and are closely related; each can be expressed in terms of the other. A set is linearly independent if and only if the family one gets when one indexes the set by itself (using its identity map) is linearly independent as a family; a family $(x_i)_{i\in I}$ is linearly independent if and only if the set $\{x_i \mid i \in I\}$ is linearly independent, *and* the indexing map $i \mapsto x_i$ is one-to-one.

P.133, material on abelian categories [<]: I suggest skipping or skimming this, since such a brief introduction is not sufficient to make the concept useful. If you want to learn something about the subject,

one old but good source is Peter Freyd's book *Abelian Categories* (Harper and Row, 1964). Rather surprisingly, Freyd's development of the theory shows that the abelian group structure that Lang and many other modern presentations of the subject assume given can in fact be deduced from simpler axioms that seem to have little to do with groups.

Re §III.4. Free modules.

P.135, second line of text [~]: Delete the assumption that S is nonempty! I will say more about this in my comments on the next section.

P.135, after proof of Theorem III.4.1 [>]:

Bilinear maps on free modules.

Note that Theorem III.4.1 characterizes *free modules* (defined in the preceding line) by a universal property parallel to the universal properties of free abelian groups and free groups. Indeed, free abelian groups are equivalent to free $\mathbf{Z}$-modules. (The *construction* of free modules on an arbitrary set will be discussed on p.137.)

The universal property of free modules is an existence and uniqueness property for module homomorphisms, also known as *linear* maps. From it, one can get a similar result for *bilinear* maps, which is very useful in constructing and studying these maps, and hence in defining algebra structures on certain free modules. This result is given in the lemma below; in the second paragraph, it is generalized further to *n-linear* maps, defined in the obvious way.

Lemma III.4.c1. *Let A be a commutative ring, and L, M and N A-modules, such that L and M are free, with bases X and Y respectively. Then for any set-map $t\colon X\times Y \to N$, there is a* unique *bilinear map $\tau\colon L\times M \to N$ whose restriction to $X\times Y \subseteq L\times M$ is t.*

More generally, the analogous result is true for n-linear maps $M_1\times\ldots\times M_n \to N$, where $M_1,\ldots,M_n$ are free on bases $X_1,\ldots,X_n$ respectively.

Proof. This could be done ''by hand'', essentially as Lang proves Theorem III.4.1, but let us note that the result is also a consequence of that theorem. For each $x\in X$, let us define a map t_x of Y into N by $t_x(y) = t(x,y)$. By Theorem III.4.1, there is a unique A-module homomorphism $\tau_x\colon M \to N$ extending t_x. Now the association $x \mapsto \tau_x$ is a set-map $X \to \mathrm{Hom}_A(M,N)$, hence there exists a unique A-module homomorphism $\tau_*\colon L \to \mathrm{Hom}_A(M,N)$ which sends each $x\in X$ to τ_x. If we define $\tau(u,v) = \tau_*(u)(v)$, we see that this is A-linear in u because τ_* is linear, and is A-linear in v because τ_* takes values in $\mathrm{Hom}_A(M,N)$, hence it is A-bilinear. From the equations uniquely characterizing the maps at each step, we find that τ is the unique bilinear map taking the desired values on elements of $X\times Y$.

The result on general multilinear maps can be gotten by induction, proving the n-linear case from the $(n-1)$-linear case just as the bilinear case was proved from the linear case above. □

P.136, start of next-to-last paragraph [>]: What Lang calls a **principal** module (using the word that is common in referring to ideals) is more often called a *cyclic* module (generalizing the use of this word in group theory).

P.136, third from last line, ''and the process can be continued'' [>]: What Lang means is that one can map a free module F_1 of finite rank onto M_1, call its kernel M_2, and if this turns out also to be finitely generated, one can map a free module F_2 of finite rank onto it, etc.. Note that in the general step, one is mapping a free module F_i onto the submodule $M_i \subseteq F_{i-1}$; hence the result can be looked at as a map $F_i \to F_{i-1}$. Because the image M_i of each of these maps is the kernel of the next, what we are getting is an exact sequence

$$\ldots \to F_3 \to F_2 \to F_1 \to F \to M \to \{0\}.$$

As Lang notes on the next page, even when one can't get free modules of *finite* rank, one can always get such an exact sequence of free modules of possibly infinite ranks. This is called a *free resolution* of M,

and is the starting point of many of the fundamental constructions of homological algebra.

You might find it interesting to look for a case where M is finitely generated, but where, taking F free of finite rank, one gets an M_1, and hence an F_1, that is not finitely generated.

P.137, second paragraph [=]: Lang already defined ''free module'' on p.135 (just before Theorem III.4.1). What he gives here is a way of constructing such a module with basis indexed by any set S. This construction can be motivated using the general heuristics for constructing universal objects that I sketched in my comment to Lang, p.74, bottom. Namely, consider an arbitrary A-module M given with an S-tuple of elements, $(f(s))_{s\in S}$. What elements of M can we name, in terms of these? Clearly, we can name all linear combinations $\Sigma\, a_s f(s)$ where the elements a_s lie in A and all but finitely many of these are zero. Moreover, given two expressions of this form, we can write down the expression of the same form which will represent the sum of the indicated elements of M, and given an expression of this form and an element $a\in A$, we can write down the expression representing the result of multiplying the indicated element of M by a. Now for particular modules M and maps f, certain distinct expressions of this form may describe the same element; but we can't see any equalities that must hold in *all* such cases; so let us take the set of all ''formal expressions'' of this sort, and define addition and scalar multiplication on these in the manner that we have seen works for actual expressions for elements of an actual A-module. We find that the result in fact forms an A-module, which we can describe as a ''module of formal expressions''. Since each such expression is uniquely determined by the S-tuple of coefficients $(a_s)_{s\in S}$, in which we have noted that almost all a_s are zero, we may more simply define our module to consist of all such S-tuples $(a_s)_{s\in S}$. As in Lang, one easily verifies that the set of S-tuples with a 1 in one place and 0's everywhere else forms a *basis*, in bijective correspondence with the given set S, so by Theorem III.4.1, we do indeed have a free module on S.

In our discussion of *free abelian groups*, we saw that all bases had the same cardinality. Is the same true for bases of modules over arbitrary rings? You might think about this. I will discuss the answer in my comments on §III.5.

P.137, discussion of projective modules [<]:

In studying free abelian groups, we noted that Lemma I.7.2 said that every surjective homomorphism of an abelian group to a free abelian group split. The same argument shows that every surjective homomorphism from an A-module to a free A-module splits, and this fact is an important tool in module theory. However, over many rings, there are modules other than free modules for which this also holds, and these may be considered ''as good as'' free modules for many purposes. They are given the name *projective modules*, and introduced here.

P.138, before **Examples** [=]: Recall that abelian groups are essentially the same as $\mathbf{Z}$-modules. Hence, let us ask, ''What are the projective $\mathbf{Z}$-modules?'', and look at this as a question about abelian groups. From condition **P 3**, we see that every such module must *embed* in a free $\mathbf{Z}$-module. But we know that every subgroup of a free abelian group is free; so every projective $\mathbf{Z}$-module is free. Since we began by observing that projectivity was a property of free modules, we can conclude that the projective abelian groups are precisely the free abelian groups. In the next paragraph, Lang will mention several other classes of rings for which all projective modules are known to be free, including two large classes for which he will soon prove this, the fields and the principal ideal domains. However, for those of you who are interested, let me give a few examples of

Non-free projective modules (optional for my Math 250A).

First, a simple-minded class of examples. If A and B are any two rings, one can form the product ring $A\times B$. It is not hard to show that any module over $A\times B$ has the form $M\times N$, where M is an A-module, N is a B-module, and the product is made an $A\times B$-module in the obvious way. In particular, the free module of rank r over $A\times B$ has the form $A^r\times B^r$. It is easy to see that a module of the form $A^m\times B^n$ will be a direct summand in the free module $A^r\times B^r$ for any $r \geq \max(m,n)$; but in general, this will not be free; for instance, it will not be free if $A = B = \mathbf{Z}$ and $m \neq n$.

Now for some more interesting examples, though these require some geometric arguments.

Let A be the ring of all continuous real-valued functions on the unit circle $x^2 + y^2 = 1$. Let us think of this circle as lying in 3-dimensional space, and let F be the set of all functions that associate to each point of that circle a vector in 3-space perpendicular to the circle at that point, in a continuous fashion. Since we can multiply such a continuous vector-valued function by a continuous scalar-valued function, F becomes a module over A. In fact, we can see that F is free of rank 2, with one generator given by the function associating to every point of the circle the unit vector in the upwards direction, and the other associating to every point the unit vector in the outward normal direction. Now comes the geometry: Let us draw a ''Möbius strip'' containing this circle, let P_1 be the submodule of F consisting of those elements whose value at every point lies in (the plane determined at that point by) the Möbius strip, and let P_2 be the submodule consisting of elements whose value at every point is perpendicular to the Möbius strip. Then we see that $F = P_1 \oplus P_2$, so that P_1 and P_2 are projective modules over A. But I claim that they are not free. To show this, we first note that if they were free, they would each have to be free on a single generator. For given two elements of one of these modules, one can construct a nontrivial linear relation that they satisfy, with coefficients in A; though the argument requires a bit of (undergraduate) analysis, and I will not go into it here. On the other hand, I claim that neither P_1 nor P_2 can be generated by a single element. For the 180° twist in the construction of the Möbius strip has the consequence that every element x of either of these modules has a point where its value is the zero vector (draw a picture!); and if we take a member y of the given module which is nonzero at that point, we see that y cannot lie in the submodule Ax. So these projective modules can neither be free on one generator, nor free on more than one generator; so they are nonfree.

The above ring A is messy from an algebraic point of view, since it contains many zero-divisors. However, with a little more care one can show that this example can actually be realized over the subring of *polynomial* functions on the circle, i.e., functions which are the restrictions to the circle of polynomial functions on the plane. This ring, $\mathbf{R}[X, Y]/(X^2 + Y^2 - 1)$, can be shown to be an integral domain. (In fact, it is isomorphic to the ring of trigonometric functions on the line generated by the functions $\sin\theta$ and $\cos\theta$. That one can express the ''half-twist'' of the Möbius strip in terms of these functions is a consequence of the fact that $\tan(\theta/2)$ can be written as a rational function in $\sin\theta$ and $\cos\theta$.)

Another geometric example: Suppose we let A be the ring of continuous real-valued functions on the 2-sphere $x^2 + y^2 + z^2 = 1$, and F the A-module of all functions associating in a continuous fashion to each point of the 2-sphere an arbitrary vector in 3-space (which you should picture as drawn at that point of the sphere). Then F is free of rank 3, with a basis consisting of the functions which associate to every point of the sphere the unit vector in the x-direction, respectively the y-direction, respectively the z-direction. Now we can decompose F as $P_1 \oplus P_2$, where this time P_1 denotes the set of functions which are everywhere *tangent* to the sphere, and P_2 the set of functions everywhere *normal* to the sphere. Here, P_2 is free of rank 1, with generator the everywhere-outward-unit-normal function. One can show that if P_1 were free, it would have to be free of rank 2, and that a basis would have to consist of two elements, whose values at each point gave a basis of the tangent plane to the sphere at that point. But this would mean that each of these functions would have the property of associating to every point of the sphere a nonzero tangent vector, in a continuous fashion. Now it is a well-known result of algebraic topology that there is no such function (''You can't comb a hairy sphere''). Hence P_1 cannot be free. In this example, too, the ring of all continuous functions can be replaced by the integral domain of all polynomial functions.

Finally, one quick noncommutative example. Let R be any ring, n a positive integer, and A the ring of $n \times n$ matrices over R. Then the set P of column vectors over R of height n can be considered a left A-module. Note that if we regard A as a left module over itself, multiplication by each member of A acts separately on the n columns of members of A. From this we can see that A is isomorphic to the direct sum of n copies of P. Hence P is projective; intuitively, one may think of it as a ''free A-module of rank $1/n$''.

P.138, **Examples**, through end of section on next page [<]: You may consider that material optional reading. If you do read it, note that at several points (once in each paragraph), Lang uses ''finite'' in the sense ''finitely generated'' which he introduced on p.129, but which I have advised you not to use.

What do we get if we apply the ideas of the second and third paragraphs of this discussion to the geometric examples of nonfree projective modules we have just sketched? It is easy to see that in the ''Möbius strip'' case, P_1 and P_2 are isomorphic to each other. Calling their common isomorphism class p, and writing f for the isomorphism class of the free module of rank 1, it can be proved that all projective modules over this ring are direct sums of copies of these two modules, and that $K(A)$ is the abelian group presented by two generators and one relation, $\langle p, f \mid 2p = 2f \rangle$. (Being a finitely generated abelian group, this must be decomposable as a direct sum of cyclic subgroups. Can you find this decomposition?) We get $K_0(A)$ by dividing out by the subgroup generated by f; the result is $\langle p \mid 2p = 0 \rangle \cong Z_2$. In the ''hairy sphere'' example, let me just note that because the direct sum of P_1 with the free module P_2 gives the free module F, the image in $K(A)$ (and hence in $K_0(A)$) of P_1 is the same as that of the free group of rank 2, even though P_1 is not itself free. In the $n \times n$ matrix ring example, if k is a field, one can show (cf. Exercise III.5:1) that $K(A)$ is infinite cyclic with generator $[P]$, and $K_0(A)$ is cyclic of order n.

Lang gives a second definition of $K(A)$ in the very last paragraph. It can be shown that this agrees with the first definition for a large class of rings A, but we can't go into this here. (There is an obvious homomorphism from the first group to the second; the nonobvious result is that for ''good'' A, this is an isomorphism.)

When he gives the second definition of $K(A)$, Lang notes that by modifying the class of modules considered, one gets still other groups. This is true, but one has to be careful. If, for instance, one starts with the class of countably generated A-modules, then one finds that for every module M in this class, the direct sum N of countably many copies of M is also in the class. But $N \cong N \oplus M$, hence M has zero image in this Grothendieck group; hence that group is trivial. Another example of how changing the class of groups can erase structure: If we take $A = \mathbf{Z}$, and consider the class of finite abelian groups, we find that the corresponding Grothendieck group is free of infinite rank, on the generators $[Z_p]$, where p ranges over the primes. On the other hand, in the Grothendieck group obtained from the class of finitely *generated* abelian groups, we have $[Z_p] = 0$, in view of the short exact sequence $0 \to \mathbf{Z} \to \mathbf{Z} \to Z_p \to 0$. From this it is easy to see that all finite groups have zero image in this Grothendieck group, and that it is, in fact, free of rank 1 on the generator $[\mathbf{Z}]$.

P.139, end of §III.4 [>]. *Further notes.* We can *dualize* the definition of projectivity. One calls a module I *injective* if every one-to-one map from I to a module M splits (i.e., is a split injective homomorphism, as defined in the comments on §I.7). This is equivalent to saying that whenever I occurs as a submodule of a module M, it is a direct summand in M. Free modules are not in general injective; e.g., taking $A = \mathbf{Z}$, we note that the free abelian group $\mathbf{Z}$ is a submodule of $\frac{1}{2}\mathbf{Z}$, but not a direct summand therein. Two examples of injective abelian groups (though we will not prove this) are $\mathbf{Q}$ and $\mathbf{Q}/\mathbf{Z}$. Although free modules are usually not injective, it is not hard to prove from Lemma I.8.c1(b) that for $A = \mathbf{Z}/p^n\mathbf{Z}$, free A-modules *are* injective, and from the results of the next section (or undergraduate linear algebra) that over a field, every module is injective. Basic results on injective modules over general rings are developed in Lang, §XX.4, which can be read immediately after the present section. This material is sometimes covered in 250B.

Recall that in §III.2, we saw that covariant and contravariant hom-functors $\mathrm{Hom}(X, -)$ and $\mathrm{Hom}(-, Y)$ on module categories do not generally preserve exactness of (short) exact sequences. In fact, the first functor preserves exactness of all such sequences if and only if X is a projective module (Lang's **P 4**), and the second, if and only Y is an injective module. (This is not hard to prove; Lang gets it as **I 2** on p.783.)

Re §III.5. Vector spaces.

P.139, first line of §III.5 [>]: More generally, a module over a *division ring* is called a vector space, and you will find that all the arguments used in this section work for vector spaces over division rings; so read them under that more general assumption. There *are* important results in the theory of vector spaces over fields that do not hold in this more general context; but these concern determinants of linear transformations, eigenvectors, and similar subjects, not treated in this section.

P.139, statement of Theorem III.5.1 [=]: Lang requires $V \neq 0$ in order to be consistent with his previous definition that a basis must be nonempty. Since we are not requiring this, you should drop that condition from this theorem.

P.140, Proof of Theorem III.5.2 [=]: The argument becomes simpler and more elegant if one takes as the key statement that if $v_1, \dots, v_m$ span V and $u_1, \dots, u_n$ are linearly independent, then $n \leq m$. Then one can drop the last seven lines on this page, in which Lang shows that the linear independence condition he is (unnecessarily) assuming on the v's carries over to the modified sequence. (Likewise, on the line after the first display on the next page, one would weaken ''is a basis of'' to ''spans''.) When this inductive argument is complete, one can conclude that if two finite subsets are *both* linearly independent and spanning sets, then their cardinalities must be equal.

P.141, second paragraph [=]: Lang leaves the infinite-basis case ''to the reader''. This is proved by the same method as for free abelian groups. You should verify that the lemma on generating sets of groups given above in the commentary to pp.41-42 also holds for modules over any ring, by the same proof. (The proper statement of that lemma is as a general result of universal algebra.) It follows that if a module is free with bases X and Y, and one of these is infinite, then they have the same cardinality.

P.141, beginning of second line of proof of Theorem III.5.3 [=]: The canonical map Lang is referring to is the map $V \to V/W$.

P.142, end of §III.5 [>]. Now read the items below.

Trouble with trivia.

Recall that when Lang defined a ''basis'' S of an A-module M on p.135, he required that S be *nonempty*. I think most mathematicians would not make such a restriction. When this condition is deleted, one finds that the empty set is a basis of the *zero module* $\{0\}$. (In verifying this, we must be careful and note that when S is empty, the phrase ''linear combination of members of S'' is not vacuous – 0 is a linear combination of members of $\emptyset$, since the empty sum is understood to be 0.) Thus, one does not have to give the statement that the zero module is free as a special proviso in the definition, as Lang does (p.135, last four words before Theorem III.4.1).

Does it matter whether we allow empty bases? Yes! If we do not, then arguments about free modules will continually require consideration of special cases. For example, under Lang's definition, his proof of Theorem III.5.3 (p.141) is incomplete: He fails to note that if f is one-to-one Ker f will ''not have a basis'', and likewise that if f is the zero map Im f ''has no basis''. Of course, the proof of the theorem in these special cases is easy, but it's nicer not to have to deal with them. The difference is all the greater in more complicated proofs, involving, for instance, infinite families of vector spaces, any of which might be zero. Moreover, if we prove a result like Theorem III.5.3 by giving separate arguments for the case where f is one-to-one or zero, this suggests that it may just be a coincidence that the same result holds in those cases as in the general case. A proof based on the correct definition makes it clear that it is not.

I think the reason Lang and some other textbook-writers specify that bases must be nonempty is the awareness that if they do not, many students will have trouble figuring out how the definition applies to empty linear combinations. But I think the best approach is for a textbook to give the *definition* without special treatment of the zero module, then to *point out*, after the definition, how this makes $\emptyset$ a basis of the zero module; and, finally, to advise any readers who have difficulty with this case to take the statement that the empty set is a basis of the zero module as an addendum to the definition, for the time being.

(Curiously, in discussing free groups in Chapter I, Lang does not require his free generating sets to be

nonempty, though he does in the case of free abelian groups.)

Free modules with and without unique rank.

We can now address the question mentioned earlier, of whether a free module over a ring R can have bases of two different cardinalities. By our comment on Lang's p.141, this is only possible if both cardinalities are finite.

To understand the question computationally, note that a free left R-module of rank n is isomorphic to the ''standard'' free module R^n. Let us write the elements of the canonical basis of R^n, (1,0,0,...), (0,1,0,...), ... (0,...,0,1), as $e_1, \dots, e_n$. A module homomorphism $f: R^m \to R^n$ is uniquely determined by the images of the m elements $e_i \in R^m$, which can be any elements $f(e_i) = \Sigma_{j \le n}\ a_{ij} e_j$ $(a_{ij} \in R)$. Note that when we apply this homomorphism f to a general element $\Sigma_{i \le m}\ c_i e_i \in R^m$, we get $\Sigma_{i,j}\ c_i a_{ij} e_j$, and computationally, the effect is that of multiplying a row vector $(c_1, \dots, c_m)$ on the right by an $m \times n$ matrix $((a_{ij}))$. (Cf. the ring theorist's principle that homomorphisms of left modules should be written on the right of elements of these modules.) It is easy to verify that composition of homomorphisms among such free modules corresponds to multiplication of matrices over R. We deduce

Lemma III.5.c1. *Let R be a ring, and m, n positive integers. Then the following conditions are equivalent:*

(i) *The free left module of rank m is isomorphic to the free left module of rank n.*

(ii) *There exist an $n \times m$ matrix A and an $m \times n$ matrix B such that*

$$AB = I_n, \quad BA = I_m. \ \square$$

Now given any homomorphism of rings, $f: R \to S$, we can apply f to the entries of any family of matrices over R, and get matrices over S that satisfy any matrix relations that our original matrices did. From this observation we get

Corollary III.5.c2. (a) *If $R \to S$ is a homomorphism of rings, and m, n positive integers such that $R^m \cong R^n$ as R-modules, then $S^m \cong S^n$ as S-modules.*

(b) *If R is a ring which can be mapped homomorphically into a division ring, then any two bases of a free R-module have the same cardinality.*

In particular

(c) *If R is a nonzero commutative ring, then any two bases of a free R-module have the same cardinality.*

Proof. We get (a) by applying the given homomorphism to the entries of matrices as in point (ii) of the preceding lemma.

To see (b), let X and Y be two bases of a free R-module F. We have already seen that if either is infinite, then they have the same cardinality, so assume both finite; say X has cardinality m and Y has cardinality n. Since any two free modules on the same number of generators are isomorphic, we have $R^m \cong F \cong R^n$. Now if R has a homomorphism to a division ring D, then by part (a), this gives a D-vector-space isomorphism $D^m \cong D^n$. But by Theorem III.5.2 (which we have noted is valid for vector spaces over division algebras), this implies $m = n$, as claimed.

Finally, if R is a nonzero commutative ring, then letting I be a maximal ideal of R, we know that R/I is a field; so we can apply (b) to the canonical homomorphism $R \to R/I$ and conclude that bases of free R-modules have unique rank. $\square$

Now the result we just used, that for any maximal ideal I of a commutative ring, R/I is a field, does *not* extend to say that for any maximal ideal I of any ring, R/I is a division ring. Thus, though the results in Lang are as true for vector spaces over general division rings as for vector spaces over fields,

their consequences for commutative rings are stronger than for general rings! Indeed, we shall now show that there exist noncommutative rings having free modules with nonunique rank.

To do this, we must first generalize our observation about writing homomorphisms among free modules as matrices. Note that by the universal properties of products and coproducts in any category, a morphism from a coproduct to a product, $a: \coprod_I X_i \to \prod_J Y_j$, is uniquely determined by the $I\times J$-tuple of morphisms a_{ij} gotten by composing a on the right with the universal morphisms (coprojections) from the X_i to $\coprod_I X_i$, and on the left with the universal morphisms (projections) from $\prod_J Y_j$ to the Y_j. In particular, since products and coproducts of finite families of modules are both direct sums, in the module case this gives a description of morphisms $a: \oplus_I M_i \to \oplus_J N_j$ as ''matrices'' of morphisms $a_{ij}: M_i \to N_j$; and it is easy to verify that these compose by formal composition of matrices. (This description uses the fact that module maps can be added as well as composed.) We can now prove

Lemma III.5.c3. (a) *Let A be any ring, M a left A-module, and m and n positive integers such that $M^m \cong M^n$ as left A-modules. Then, letting R denote the ring $\mathrm{End}_A(M)$, we have $R^m \cong R^n$ as left R-modules.*

In particular, as a concrete class of examples

(b) *If R is the full endomorphism ring of an infinite-dimensional vector space V, then $R^m \cong R^n$ as left R-modules for all positive integers m and n.*

Proof. (a): The given isomorphism $M^m \cong M^n$ will be described by an $m\times n$ matrix and an $n\times m$ matrix of module homomorphisms $M \to M$, i.e., elements of R, which compose to the identity matrices. Hence by Lemma III.5.5, above $R^m \cong R^n$ as left R-modules.

(b): Since any infinite cardinal α satisfies $m\alpha = n\alpha$ for all positive integers m and n, we see (by working with bijections of bases) that infinite-dimensional vector spaces V satisfy $V^m \cong V^n$, from which the assertion follows by part (a). (We could equally well have stated this result for endomorphism rings of free modules of infinite rank over an arbitrary ring.) □

The above result was easy to prove. There are more delicate results that are harder, such as that there exist rings R such that all R-modules of *even* finite positive dimension are mutually isomorphic, but are not isomorphic to free modules of *odd* dimension. Cf. P. M. Cohn, *Some remarks on the invariant basis property*, Topology, **5** (1966) 215-218.

The functorial approach (sketch, optional). You may wonder why matrices were needed to get the above results. There is in fact an approach I think more enlightening, though it involves some material we have not developed. If $f: R \to S$ is a ring homomorphism, then one can regard any left S-module M as a left R-module by defining $r\cdot x = f(r)x$ $(r\in R,\ x\in M)$. The resulting functor S-**Module** $\to$ R-**Module**, called ''restriction of scalars from S to R'', has a left adjoint, called ''extension of scalars from R to S''. (Lang introduces this construction in §XVI.4, though only for R and S commutative.) Now note that the composite of the above restriction of scalars functor with the forgetful functor R-**Module** $\to$ **Set** is the forgetful functor S-**Module** $\to$ **Set**. Since adjoints of composite functors are composites of adjoint functors, with order of composition reversed, we see that the free R-module functor **Set** $\to$ R-**Module**, followed by the extension of scalars functor R-**Module** $\to$ S-**Module**, gives the free S-module functor **Set** $\to$ S-**Module**. In particular, the extension of scalars functor takes the free R-module of any rank n to the free S-module of rank n. Hence if there is an isomorphism between the free R-modules of ranks m and n, there will be an isomorphism between the free S-modules of ranks m and n, as described in Corollary III.5.6(a) above. The trick of applying the ring homomorphism $R \to S$ to entries of matrices simply mimics the way this functor acts on morphisms among free modules.

The non-unique-ranks construction of Lemma III.5.c3(b) also has a functorial interpretation, this time not even requiring adjoint functors. Observe that a left A-module M has a natural right $\mathrm{End}({}_AM)$-module

structure. Using this, we can make $\mathrm{Hom}_A(M,N)$ into a left $\mathrm{End}({}_AM)$-module for each left module N, in a functorial manner. Clearly, when $N = M$, this module is the free module of rank 1, from which it can be deduced that, writing $R = \mathrm{End}({}_AM)$, the functor $\mathrm{Hom}_A(M,-)$ takes M^n to the free left R-module of rank n. Hence if $M^m = M^n$ as A-modules, we get $R^m = R^n$ as R-modules.

Incidentally, the trick used in showing that over a commutative ring, free modules of different ranks were nonisomorphic is an instance of a method common in modern mathematics: *To show objects are nonisomorphic, find a functor taking them to objects you can already prove nonisomorphic.* We used this earlier to show free abelian groups of distinct ranks nonisomorphic. Though we did not have the theory of dimension of vector spaces available at the time, we saw that the groups one got by applying to such groups the functor $A \mapsto A/pA$ could be distinguished by their *orders*. Another standard use of the same method is in showing that the free (nonabelian) group of ranks m and n cannot be isomorphic unless $m = n$ (which was actually never shown in §I.12). To do this one applies the *abelianization* functor $G \mapsto G/G^c$. It is easy to verify by universal properties (or by Lemma II.3.c3) that the abelianization of a free group of rank n is a free abelian group of rank n; hence, since we have shown that free abelian groups have unique rank, the same is true of free groups.

Let me mention an interesting pair of open questions on bases of vector spaces.

Open Questions III.5.c4. *Let k be a field, n a positive integer, and V an n-dimensional vector space over k.*

(a) (J. Kahn) *Suppose we are given n^2 bases $\mathfrak{B}_{11}, \mathfrak{B}_{12}, \dots, \mathfrak{B}_{nn}$ of V (not necessarily distinct. Think of these as written in the cells of an $n \times n$ array.) Is it possible to choose from each $\mathfrak{B}_{ij}$ an element b_{ij}, in such a way that for each i, $\{b_{ij} \mid j = 1, \dots, n\}$ is a basis of V, and for each j, $\{b_{ij} \mid i = 1, \dots, n\}$ is a basis of V? (I.e., can we choose one element from each cell so that in each row, and in each column, the chosen elements form a basis of V?)*

(b) (G.-C. Rota) *Same question, under the additional assumption that for each i, all the $\mathfrak{B}_{ij}$ are the same. (So this question asks whether, given n bases $\mathfrak{B}_i$ of V, we can recombine them into n bases $\mathfrak{B}'_j$, each using one element of each $\mathfrak{B}_i$, so that each element of each $\mathfrak{B}_i$ is used in only one $\mathfrak{B}'_j$.)*

(See R. Huang and G.-C. Rota, *On the relations of various conjectures on Latin squares and straightening coefficients*, Discrete Math., **128** (1994) 225-236.)

Re §III.6. The dual space and dual module.

P.142, first paragraph of §III.6 [>]: Note that the duality $E^\vee$ studied in this section is quite different from the duality $E^\wedge$ of §I.9. In the last sentence of this paragraph, Lang makes two claims which he does not justify: that the map $E \to E^{\vee\vee}$ is always injective, and that it is not always surjective. I will show both facts in the comment to be read following the proof of Theorem III.6.1.

P.143, statement of Theorem III.6.1 [=]: In the first line, delete the word ''finite''. (Lang is using the word in the sense ''finitely generated'', but it is redundant in view of the next line.) In the third line, before ''f_i *is*'', add the words ''*for each* i''.

P.143, first display [=]: Add, on the right, ''$= a_1 c_1 + \dots + a_n c_n$''. Also, in the line after the second display, ''K'' should be ''A''.

P.143, end of proof of Theorem III.6.1 [>]. We can now show the two facts asserted at the bottom of p.142. Let us write $\theta\colon E \to E^{\vee\vee}$ for the map described there, let $(x_i)_{i \in I}$ be a basis of E, and let (f_i) be defined as in Theorem III.6.1.

To show θ is injective, consider an element $x \in E$, written in terms of the x_i. If $x \neq 0$, there will be some $i \in I$ such that x_i occurs with nonzero coefficient a_i in this expression. Thus, $\theta(x)(f_i) = f_i(x) = a_i \neq 0$, showing that $\theta(x) \neq 0$.

On the other hand, I claim that when A is a field and E is infinite-dimensional, θ is not surjective. We first note that the argument used in the proof of Theorem III.6.1 still shows that the family (f_i) is

linearly independent (though the proof that if E has finite rank that set generates $E^\vee$, which involves taking a summation over all i, does not work in the non-finitely-generated case). Hence we can enlarge this family to a basis $\mathcal{B}$ of the vector space $E^\vee$. Let us now define $\varphi \in E^{\vee\vee}$ to be the unique linear map $E^\vee \to A$ which takes all elements of $\mathcal{B}$ to 1. If φ were of the form $\theta(x)$ $(x \in E)$, it would take each f_i to the coefficient of the basis element x_i in the expression for x, so it would take on nonzero values at only finitely many of the elements f_i. Thus, φ cannot have the form $\theta(x)$, so θ is not surjective.

Note that for any base ring A, if E is free on a basis I, then $E^\vee$ is isomorphic to an I-fold direct *product* of copies of A. Thus, our old Exercise I.7:1 effectively asked you to show that for E free, $E^\vee$ may fail to be free.

P.143, statement and proof of Corollary III.6.2 [=]: In the second line of the statement, ''V'' should be ''E''. The proof is sloppily worded. Again writing $\theta\colon E \to E^{\vee\vee}$ for the map referred to in the corollary, one should replace $\{x_1, \ldots, x_n\}$ in the second line by $\{\theta(x_1), \ldots, \theta(x_n)\}$, and ''$E = E^{\vee\vee}$'' by ''$\theta\colon E \to E^{\vee\vee}$ is an isomorphism''.

P.143, Theorem III.6.3 [=]: The assumption that U, V and W are free is much stronger than we need; we only need an assumption on U, and this assumption is that it be projective (which is weaker than free). From that condition, one sees that the given exact sequence splits, and, as we noted in our comments on split exact sequences under §I.7, this makes the sequence a ''direct sum'' of the two exact sequences $0 \to W \to W \to 0 \to 0$ and $0 \to 0 \to U \to U \to 0$. It is not hard to see that $\mathrm{Hom}(-, A)$ preserves direct sums, and therefore preserves exactness of such sequences, proving the theorem. In fact, this argument shows that the conclusion can also be strengthened, replacing $\mathrm{Hom}(-, A)$ by $\mathrm{Hom}(-, M)$ for any module M.

P.145, **Remark 2** [<]: Let me precede this with some remarks of my own.

The context in which Lang introduces this duality is unnecessarily restricted. First, the functor $\mathrm{Hom}(-, A)$ can be applied to arbitrary modules, not just free modules. (Indeed, in the last sentence of the first paragraph, Lang applies $^\vee$ to $E^\vee$, which may not be free if E is free of infinite rank.) Although this functor is ''uninteresting'' when applied to certain non-free modules (e.g., when $A = \mathbf{Z}$ and E is either a torsion abelian group or $\mathbf{Q}$, then $E^\vee = 0$), there are other non-free modules for which it is important. When it is applied to finitely generated *projective modules*, one can show that the analogs of all the theorems of this section hold. You might also find it interesting to examine the case where $A = k[X, Y]$, a polynomial ring over a field, and E is the ideal (X, Y); in particular, to compute $E^\vee$, $E^{\vee\vee}$, and the natural map $\theta\colon E \to E^{\vee\vee}$.

Lang has taken the base-ring A to be commutative. If this is not assumed, and E is a left A-module, it turns out that the abelian group of left A-module homomorphisms $E^\vee = \mathrm{Hom}(E, A)$ has a natural structure of *right* A-module; and likewise, the abelian group of right-module homomorphisms from a right A-module into A has a structure of *left* A-module. (More details in my comments in §XIII.1.) Thus, one gets a duality between right and left modules, which, when these modules are finitely generated and free (or projective), still has all the properties Lang proves in the commutative situation.

In **Remark 2** on this page, Lang notes a different kind of duality, taking an R-module E to the set $E^\wedge = \mathrm{Hom}(E, \mathbf{Q}/\mathbf{Z})$ of *abelian group* homomorphisms $E \to \mathbf{Q}/\mathbf{Z}$. What he fails to mention is that this set can be made an R-module in a natural way! For instance, if E is a left R-module; then $E^\wedge$ becomes a right R-module by defining, for each $f \in E^\wedge$ and $r \in R$, the element $fr \in E^\wedge$ to be the map that acts by $fr(x) = f(rx)$. Here too, we get a natural map $E \to E^{\wedge\wedge}$. This is always one-to-one, but rarely onto. (It is onto when E is finite in the set-theoretic sense, by the results of §I.9.)

For this duality by $\mathbf{Q}/\mathbf{Z}$, and also for duality of vector spaces by the construction of this section, the failure of surjectivity of the embedding of an object E in its double dual can be considered to result from the existence of ''strange'' functionals on the dual. In some contexts it is possible to exclude these by introducing a topology on the dual, and restricting attention to continuous functionals. When one does this for the $E^\wedge$ construction, it is also natural to replace $\mathbf{Q}/\mathbf{Z}$ by $\mathbf{R}/\mathbf{Z}$. For any topological abelian group

E one then defines $E^\wedge$ to consist of all *continuous* homomorphisms $E \to \mathbf{R}/\mathbf{Z}$, one puts a natural topology on this group, and one shows that if E is *locally compact*, then so is $E^\wedge$, and the natural map $E \to E^{\wedge\wedge}$ is then an isomorphism of topological groups. (This is Pontrjagin duality, mentioned by Lang in the next paragraph; cf. Walter Rudin *Fourier Analysis on Groups*, Interscience, 1962.) One finds, inter alia, that under this duality, the compact abelian groups are the duals of the discrete abelian groups; for example, this duality takes the short exact sequence $0 \to \mathbf{Z} \to \mathbf{R} \to \mathbf{R}/\mathbf{Z} \to 0$ to itself, interchanging the right- and left-hand objects. Similarly, for a vector space E over an arbitrary field k, one can put a natural topology on $E^\vee = \mathrm{Hom}(E, k)$, such that for ''good'' E, the natural map $E \to E^{\vee\vee}$ is an isomorphism of topological vector spaces. In particular, discrete spaces are ''good'', and are dual to what are called ''linearly compact'' spaces. This is ''Lefschetz duality'', but is not so well known.

Re §III.7. Modules over principal rings.

I motivated the structure theory of finitely generated abelian groups in my comments on §I.8. Since the theory of modules over principal ideal domains closely parallels that theory, I will not repeat that discussion here. Though Lang generally leaves the non-finitely-generated case to the reader, or refers to an Appendix, it is treated in detail in those comments.

P.147, Corollary III.7.2 [>]: Lang indicates two ways of proving this. Actually, these yield different results, each stronger than the result he states. The proof he writes out shows that if E is a module over a PID which can be generated by n elements, then any submodule of E can be generated by $\leq n$ elements. The other proof shows that any submodule of a finitely generated module over any *Noetherian ring* is finitely generated (but not necessarily by fewer generators than the given module. E.g., if R is the polynomial ring $k[x, y]$ in two indeterminates over a field, and $E = R$, the free module of rank 1, then the submodule generated by x and y cannot be generated by fewer than two elements, the submodule generated by x^2, xy and y^2 cannot be generated by fewer than three elements, etc..)

P.148, first paragraph [~]: This argument shows that any finitely generated torsion-free module M over an integral domain R embeds in a free module of finite rank. When R is, in fact, a principal ideal domain, it follows that M is free, which is the fact he will use next.

P.148, Lemma III.7.4 [~]: Lang is really re-proving the fact that a free module satisfies condition **P2** (p.137), i.e., is projective.

P.149, bottom, first sentence of proof [<]: If we take a finite generating set for E, we may choose a nonzero exponent for each member of this set, and take a common multiple $a \in R$ of these elements. Thus a will annihilate all members of our generating set, hence all members of E, justifying this sentence.

P.151, statement of Theorem III.7.7 [>]: Note where Lang above the display refers to the $R/(q_i)$ as nonzero, this is equivalent to saying that the q_i are nonunits.

P.151, next-to-last line of text [=]: After ''$i = 1, \ldots, l$'' add, '', with some of these rows 'padded' with zeros on the left, if necessary, so that they all have the same length r as the longest row''.

P.153, Theorem III.7.8 [<]: Any finitely generated R-module E can be written as a homomorphic image of a free module F of finite rank, hence as a factor-module of F by a submodule M. We have just shown that E can be written as a direct sum of cyclic modules $R/(a_i R)$. Can we write F as a direct sum of free modules of rank 1, Rx_i, so that M is the sum of the submodules $Ra_i x_i$; i.e., can we ''realize'' the decomposition of Theorem III.7.7 via the given presentation of E? A positive answer is given by the next result, which is thus a strengthening of Theorem III.7.7.

Incidentally, in the statement of the theorem, the condition $M \neq 0$ is pointless. The result is clearly true for $M = 0$, and the induction with which the proof ends on the next page implicitly uses that fact, since M_1 may be 0.

Pp.153-154, sentence on bottom of one page and top of next [=]: To prove this assertion, note that because R is a principal ideal domain, the ideal $(g(x_1)) + (\lambda_1(x_1))$ can be generated by one element, say

$$(ug(x_1) + v\lambda_1(x_1)) \;=\; (g(x_1)) + (\lambda_1(x_1)).$$

But this implies

$$(ug + v\lambda_1)(M) \;\supseteq\; ((ug + v\lambda_1)(x_1)) \;=\; (g(x_1)) + (\lambda_1(x_1)) \;\supseteq\; (\lambda_1(x_1)) \;=\; \lambda_1(M)$$

(where the middle step is the preceding display). Now by maximality of $\lambda_1(M)$, the two inclusions must be equalities; the second one now shows that $g(x_1) \in (\lambda_1(x_1)) = (a_1)$, as Lang claims.

P.154, Theorem III.7.9, line 1, and preceding sentence [=]: Precisely, an ''elementary matrix'' means a matrix which differs from the identity matrix *either* by changing one off-diagonal entry 0 to a nonzero value, or by replacing one diagonal entry 1 to another unit of the ring. (Lang has a definition of the term on p.540 that leaves out the latter sort of element. I haven't checked whether that is what he really intends there; but here, elements of both sorts are needed to make the condition of generating $GL_n(R)$ plausible.)

You should verify that multiplying a matrix A on the left by an elementary matrix has the effect of a row operation on A, while multiplying on the right by an elementary has the effect of column operation. The operations of interchanging two rows and columns are not themselves of this form, but each can be expressed as the product of a small number of such operations. Thus, the assumption in the first sentence of the theorem implies that any operation of left or right multiplication by invertible matrices is equivalent to a finite number of row, respectively, column, operations.

In fact, if we delete that first sentence from the theorem, the result can still be made true by changing the phrase ''*with a finite number of row and column operations*,'' to ''*by multiplying on the right and on the left by invertible square matrices*''.

Incidentally, note that the matrix in the statement of the theorem is not assumed square.

P.155, end of §III [=]: The final paragraph is a technical remark which you can ignore.

Though the results of this section classify the structures of all finitely generated modules over PID's, if one drops the hypothesis of finite generation, one finds oneself in a very nontrivial area – even if one limits attention to torsion-free modules of finite rank (i.e., having no infinite set of linearly independent elements; equivalently, embeddable in a finite-dimensional vector space over the field of fractions of R). Such modules (over Dedekind domains, a generalization of principal ideal domains) are the subject of a book in preparation by E. Lee Lady, which can be viewed at `http://www.math.Hawaii.Edu/~lee/book`.

Re §III.8. Euler-Poincaré maps.

P.156, top, italicized condition [~]: This condition is satisfied by various natural invariants – the dimension of a finite-dimensional vector space; the order of a finite group (regarded as a member of the multiplicative group of rationals), etc.. Hence one gives a name to an arbitrary function satisfying this condition, and studies such functions in the abstract.

P.156, italicized sentence near bottom [=]: Lang hasn't said what he means by ''are equivalent''; he means ''have, up to isomorphism, the same factors, with the same multiplicities, but possibly in a different order''. The Jordan-Hölder Theorem, Theorem I.3.5, is a consequence of the Schreier Refinement Theorem, which was proved on p.22 – for groups, not modules. However, the construction in the proof used only joins, intersections, and canonical isomorphisms, all of which respect module structures if these are present. Hence if we apply this proof to towers (which Lang here calls finite filtrations) of modules, regarded as abelian groups, then the isomorphisms obtained are module isomorphisms. So the Schreier Refinement Theorem also holds for modules; hence so does the Jordan-Hölder Theorem.

Re §III.9. The Snake Lemma.

P.158, statement of **Snake Lemma** [=]: The expression $f^{-1} \circ d \circ g^{-1} z''$ gives an element of N'; what Lang means, however, is the image of that element in Coker d'. (One could say that there should be an additional map at the other end of that composite as well, the inclusion of Ker d'' in M''. However, it is common to leave inclusion maps unwritten, but not, in general, to leave factor-maps unwritten.)

Re §III.10. Direct and inverse limits. This is an important subject, begun in §I.10. As with that section, I haven't so far had time to treat this material when I have taught Math 250, but Bjorn Poonen did cover it in Fall 2003, and gave me some errata; and I present these errata and some further observations I made in checking those errata out.

P.160, definition of **directed family of morphisms** [~]: Recall that in §I.10, Lang used ''directed family'' as short for ''inversely directed family'', the structure from which we construct inverse limits. Here, however, he begins with the concept relevant to direct limits and (in agreement with common usage) calls this a ''directed family''; so the phrase can no longer be used in the sense used previously. In the bottom paragraph of p.161, beginning ''We now reverse the arrows'', the resulting system is an inversely directed family, though he does not name it here.

P.162, fourth line of **Example**, and line below last display [=]: Change ''projective limit'' to ''inverse limit''.

Chapter IV. Polynomials.

Re §IV.1. Basic properties of polynomials in one variable.

P.174, last display and preceding line [=]: The holds by the italicized result on the lower part of p.100, ''*If we assume that at least one of the leading coefficients ...*''.

P.175, end of Proof of Theorem IV.1.2 [>].

Euclidean rings.

The technique used in proving the above theorem is called the ''Euclidean algorithm'' because it goes back to Euclid's technique for finding the greatest common divisor of two positive integers $a_1 > a_2$: subtract from a_1 that multiple of a_2 that will leave a nonnegative remainder $a_3 < a_2$; verify that a_2 and a_3 will have the same greatest common divisor (i.e., will generate the same ideal) as a_1 and a_2, and, if a_3 is again positive, repeat the process with a_2 and a_3, till one eventually gets a zero term, $a_{n+1} = 0$. Then the preceding term a_n must generate the ideal $a_1\mathbf{Z} + a_2\mathbf{Z}$. If one merely wants to prove that this ideal is principal, this argument can be shortened: Let a be a least positive element of this ideal, and prove that it must be a generator.

As the proof in Lang shows, a similar argument (due to Simon Stevin, *Arithmétique*, 1585) can be applied to polynomials, using the degree-function. In fact, there is a wide class of rings in which this method can be used, characterized in

Definition IV.1.c1. *If R is an integral domain, a* Euclidean function *on R means a set-map φ from $R - \{0\}$ to the nonnegative integers, such that*

(i) *if $a, b \in R - \{0\}$, then $\varphi(ab) \geq \varphi(b)$.*

(ii) *if $a \in R$, $b \in R - \{0\}$, then there exist $q, r \in R$ such that $a = qb + r$ and either $r = 0$ or $\varphi(r) < \varphi(b)$.*

An integral domain which admits a Euclidean function is called a Euclidean domain *or* Euclidean ring.

Mimicking the proof that $\mathbf{Z}$ and $K[X]$ are PID's, one gets

Theorem IV.1.c2. *Every Euclidean domain is a PID.* □

Another ring easily shown to be Euclidean, and hence a PID, is $\mathbf{Z}[i] \subseteq \mathbf{C}$, via the norm function $\varphi(m + ni) = m^2 + n^2$. (The conclusion of condition (ii) above is equivalent to the statement that there exists a $q \in \mathbf{Z}[i]$ whose distance from a/b is < 1. To prove this, verify, more generally, that for any complex number α there is a $q \in \mathbf{Z}[i]$ whose distance from α is $\leq 1/\sqrt{2}$.)

In fact, the Euclidean algorithm is essentially the only ''easy'' way known to show that a ring is a PID. (What about using the fact that a localization $S^{-1}R$ of a PID R is a PID? Well, if R was proved a PID with the help of a Euclidean function φ, then a localization $S^{-1}R$ will also be Euclidean, via the function that takes each element x to the least of the values $\varphi(a)$ as as^{-1} runs over all expressions for x.) However, the handout ''A principal ideal domain that is not Euclidean'' shows how to get one example not of that form.

There are some generalizations of the definition of Euclidean ring, which still allow one to prove that the ring in question is a principal ideal domain. In particular, one may allow φ to have range in a general *well-ordered* set, rather than the nonnegative integers; but I don't know any cases where this generalization can be used but the standard concept cannot. Condition (i) may also be dropped from the definition; however, the class of rings admitting a function φ satisfying this weakened condition will actually be the same as the class admitting a φ that satisfies the original condition. For if φ satisfies (ii), then the function φ' defined by letting $\varphi'(a)$ be the minimum, over all nonzero $c \in R$, of $\varphi(ca)$, will satisfy both (i) and (ii).

P.175, middle, definition of **irreducible** [=]: This term was defined on p.111 (beginning of §II.5), and does not need a separate definition for polynomials. What Lang should say here is that for k a field, a polynomial $f(X)$ will be irreducible if and only if it has the stated property, since in $k[X]$, the nonzero constants are the invertible elements.

P.176, Corollary IV.1.5. [=]: Delete the final phrase ''*i.e., f induces the zero function*''. (It is true, but not the meaning of what precedes.)

P.176, next-to-last line, ''... the same function f ...'' [=]: The fact being used here is that a field k of q elements satisfies the identity $x^q = x$. This was proved in the case $k = \mathbf{Z}/p\mathbf{Z}$ in the second **Example** on p.99. You should check that the same argument works for any finite field.

P.177, Proof of Theorem IV.1.9 [<]: Actually, Lang did most of the work needed here when he proved Proposition I.4.3(vi), p.25, so he should have called on that result in this proof! By that result, it suffices to show that no U contains no more than p roots of $X^p - 1$ for any prime p; but this is immediate by the preceding theorem.

P.177, Theorem IV.1.9 [>]: Note that this shows that if $p < q$ are primes, then $\mathbf{Z}_q^*$ has at most one subgroup of order p. This is the statement which we saw, in our discussion of groups of order pq, was needed to deduce that there is no more than one isomorphism class of nonabelian groups of that order. So at last that argument is complete!

P.178, second sentence of third paragraph [>]: In fact, Lang will prove that $\mathbf{C}$ is algebraically closed in Example 5 on p.272.

P.178, third-from-last display [=]: The first formula is clear. To verify the second formula by writing f and g as sums of monomials, and computing the two sides, is tedious and unenlightening. A nicer way is to observe that the maps $(f,g) \mapsto (fg)'$ and $(f,g) \mapsto f'g + fg'$ are both *bilinear* maps of A-modules, hence they will be equal if they agree as f and g range over the spanning set of powers of X (cf. lemma in our note to p.135 of Lang). Thus, to prove that equation, it suffices to prove it in the case $f = X^m$, $g = X^n$, a quick computation.

P.179, last paragraph (excluding last partial sentence) [>]: Here properties of the pth-power map on a ring of characteristic p are obtained using the fact that certain binomial coefficients are divisible by p. In fact, this property of the binomial coefficients can be proved from a result we already know about the pth-power map on such a ring! Namely, we have seen (using properties of the group of units) that in $\mathbf{Z}/p\mathbf{Z}$, the pth power map is the identity. It follows that the polynomial $(x+1)^p - x^p - 1$ over that field, which has degree at most $p-1$, and whose coefficients are these binomial coefficients modulo p, has all p elements of this field as roots. Since the number of roots is greater than the degree, all coefficients of this polynomial must be zero.

(There are other interesting ways of showing that these coefficients are divisible by p. For instance,

the νth coefficient is the number of ν-element subsets of $\{1,...,p\}$. Now the cyclic permutation $(1,...,p)$ has order p, and if ν is neither 0 nor p, it clearly acts without fixed points on the set of ν-element subsets of $\{1,...,p\}$. Hence the number of such subsets must be divisible by p.)

Pp.179-180, sentence on bottom of one page and top of next [=]: The *Frobenius endomorphism* should be defined to be the map $x \mapsto x^p$. The maps $x \mapsto x^{p^r}$ are the powers of the Frobenius. (When one considers k-algebras A for k a field of p^r elements, one sometimes calls the map $x \mapsto x^{p^r}$ the Frobenius endomorphism of A, since it, and not $x \mapsto x^p$, is an endomorphism of A as a k-algebra. But in that usage, there is a particular power of p relevant to the situation, not an arbitrary power, as Lang's wording suggests. I will always use ''Frobenius endomorphism'' to mean the map $x \mapsto x^p$.)

P.180, end of §IV.1 [>].

The Jacobian conjecture (optional).

There is a longstanding open problem on polynomial algebras, which this is a natural time to mention.

We have seen that the operation of *differentiation* of polynomials, which we learned in calculus, can be carried over formally to a function $D\colon A[X] \to A[X]$ for any commutative ring A, which has the algebraic properties of differentiation, though it is not defined using limits. Note, incidentally, that it is not an algebra homomorphism. Like such a homomorphism, it respects the A-module structure, but instead of the law $h(st) = h(s)h(t)$, it satisfies $D(st) = D(s)t + sD(t)$. A function from an A-algebra into itself with these properties is called a *derivation* over A.

We can likewise define *partial differentiation* operators $D_X, D_Y\colon A[X,Y] \to A[X,Y]$. These are each derivations over A, and they commute with one another: $D_X D_Y = D_Y D_X$.

Let us now recall that a pair (f, g) of real-valued functions on the plane $\mathbf{R}\times\mathbf{R}$ is equivalent to a single map of the plane into itself, namely $(x,y) \mapsto (f(x,y),\ g(x,y))$, and that if f and g are differentiable, then the information about this map analogous to the information given by the derivative of a function of one variable is given by the *Jacobian matrix*,

$$\begin{pmatrix} D_X f & D_Y f \\ D_X g & D_Y g \end{pmatrix}. \tag{c24}$$

In particular, this satisfies a *chain rule* saying that the Jacobian matrix of a composite of two maps is the product of the Jacobian matrices of the two maps. (If the two maps correspond to pairs of functions (f,g) and (h,i) respectively, then their composite corresponds to $(f(h,i),\ g(h,i))$, and the statement of the chain rule is, precisely, that the Jacobian matrix of the composite map, evaluated at a point (x,y), is the product of the Jacobian matrix of (f,g) evaluated at the image of (x,y) under (h,i), and the Jacobian matrix of (h,i) evaluated at (x,y) itself.)

This suggests that we should think of a pair (f, g) of elements of $A[X,Y]$ as the ''algebraic analog'' of a map of the plane into itself, and define the Jacobian matrix of such a pair by (c24), using our algebraically defined partial differentiation operations. If we define the ''composite'' of two such pairs (f,g) and (h,i) as the pair $(f(h,i),\ g(h,i))$, it is not hard to prove the chain law for these Jacobian matrices. Note that the pair (X, Y) is a neutral element for this composition, and has for Jacobian matrix the identity matrix.

There is also an algebraic interpretation of such pairs: Any A-algebra endomorphism α of the polynomial ring $A[X,Y]$ is determined by the pair of elements $(\alpha(X),\ \alpha(Y))$, which we can denote (f,g), and conversely (by the universal property of polynomial rings), every such pair corresponds in this way to an endomorphism. Composition of endomorphisms corresponds to substitution of one pair into another, although the order of composition is the reverse of the order we get when composing maps of the plane. (This is because there is a contravariant relation between points of the plane and functions on the plane.)

Given a pair (f,g) (equivalently, an endomorphism of $A[X,Y]$), it is not clear how to tell whether it

is *invertible* under the above composition, i.e., corresponds to an A-algebra *automorphism* of $A[X, Y]$. But we can give a necessary condition: by our above observations on the chain rule, if such an endomorphism is invertible, the Jacobian matrix of the pair of functions must be invertible as a matrix of elements of $A[X, Y]$. This is equivalent to saying that the *determinant* of the Jacobian matrix is an invertible element of $A[X, Y]$.

Let us take for A a field k. Then the invertible elements of $k[X, Y]$ are the nonzero elements of k; so if an automorphism is invertible, the determinant of its Jacobian matrix will be nonzero scalar. Is the converse true? It definitely is not if k has finite characteristic p, for in this case the pair $(X^p + X, Y)$ has identity Jacobian matrix, but the corresponding endomorphism is not invertible. So assume k has characteristic 0. To prove the desired converse, it would suffice to prove it in the case where the determinant is 1, since any endomorphism of $k[X, Y]$ with nonzero scalar Jacobian determinant can be transformed into one with determinant 1 by composing with an invertible linear transformation. The problem thus takes the form

Open Question IV.1.c3 (Jacobian conjecture)**.** *Prove or disprove: If k is a field of characteristic 0, then any k-algebra endomorphism of $k[X, Y]$ whose Jacobian matrix has determinant 1 is an automorphism.*

We remark that when $k = \mathbf{R}$, the analogous statement for *real analytic* maps of the plane into itself is false. For let $f : \mathbf{R} \to (-1, 1)$ be a real analytic function with everywhere positive derivative. Then the map $(x, y) \mapsto (f(x), f'(x)^{-1}y)$ is area-preserving, hence has Jacobian determinant 1, but it carries the plane into the strip $(-1, 1) \times \mathbf{R}$, and so is not onto. If we compose this f with a locally area-preserving real analytic map that winds the strip onto an annulus, we can get an example that is also not one-to-one.

Re §IV.2. Polynomials over a factorial ring.

The goal of this and the next section is to study conditions under which polynomials can or cannot be factored, and in particular, do or do not have roots.

P.180, first paragraph of this section [>]: For an example of the concept of the ''order'' of an element at a prime, let $A = \mathbf{C}[t]$ and $p = t - a$. Then $\mathrm{ord}_p(f)$ is the order (= multiplicity) of the zero of the function f at the point a. (If f does not have a zero at a, we say ''the order of its zero there is 0''.) This is probably the source of the name. More generally, if f is a member of the field of fractions of this ring, $\mathbf{C}(t)$, called ''the field of rational functions in t'', then $\mathrm{ord}_p(f)$ is again the order of the zero of f at a if f is defined at a, while if f has a pole at a, it is the negative of the order of this pole.

In the next sentence, Lang extends this concept from elements of K to polynomials $f \in K[X]$. Note that his definition can be simplified to say that the order of $f \in A[X]$ at the prime $p \in A$ is the minimum of the orders at p of *all* the coefficients of f, zero and nonzero.

P.181, statement of Gauss's Lemma [>]:

Content, and Gauss's Lemma. To understand what the ''content'' of a polynomial $f \in K[X]$ means, think first of the case where $f \in A[X]$. In this case, it is the greatest common divisor of the coefficients of f; i.e., the ''largest'' member of A that divides f. The easiest description of the content of a general $f \in K[X]$ is gotten by bringing the coefficients of f to a common denominator, i.e., writing $f = c^{-1}g$ where $c \in A$, $g \in A[X]$; then the content of f will be c^{-1} times the greatest common divisor of the coefficients of g. (If one defines it this way, one must, of course, prove that it is unique up to multiplication by units.) Still another way of interpreting the concept is by generalizing the concept of ''divisibility'', defining $u \in K$ to ''divide'' $v \in K$ ''with respect to A'' if $v = au$ for some $a \in A$. Then the content of f is simply the greatest common divisor, in this sense, of the coefficients of f.

The concept of ''content'' is mainly a tool in the formulation and proof of Gauss's Lemma, and its consequence Theorem IV.2.3, so don't expect to see it outside this and the next section. Actually, Gauss's

Lemma is often, and, I think, most usefully, stated in a way that doesn't refer to ''content'' – though that concept, or some equivalent, is still needed in the proof. This statement, which you should verify can be deduced from the statement in Lang, is

(c25) (Gauss's Lemma, alternative formulation) If A is a UFD, with field of fractions K, and we have a polynomial $f \in A[X]$ and a factorization $f = gh$ with $g, h \in K[X]$, then there exists a nonzero element $u \in K$ such that both terms in the factorization $f = (ug)(u^{-1}h)$ lie in $A[X]$.

Re §IV.3. Criteria for irreducibility.

P.184, proof of Theorem IV.3.1 [>]: This proof can be made simpler in two ways. First, instead of reducing to the ''content 1'' case, we can use the version of Gauss's Lemma given in (c25) above to conclude that the given factorization can be taken in $A[X]$. Second, observe that if our polynomial f has factorization gh in $A[X]$, where g and h are nonunits, then since the leading terms of g and h are not divisible by p, the images of g and h in $(A/(p))[X]$ both have positive degree. But the image of their product f in that ring has the form cX^n, and the only factors of a monomial in a polynomial ring over an integral domain are monomials; hence the images of g and h are monomials, and both therefore have constant term zero. Thus, g and h have constant terms divisible by p, hence their product f has constant term divisible by p^2, contradicting the hypothesis on a_0.

Though the assumption that A is a UFD is used to reduce the proof of this theorem to showing that f has no factors of positive degree in $A[X]$, the proof of the latter fact, which is the main part of the argument, does not require that A be a UFD. Thus, we get

(c26) (Eisenstein's irreducibility theorem, variant form.) Let A be an integral domain, $f(X) = a_n X^n + \ldots + a_0 \in A[X]$, and $\mathbf{p}$ a prime ideal of A, such that $a_n \notin \mathbf{p}$, $a_i \in \mathbf{p}$ for all $i < n$, and $a_0 \notin \mathbf{p}^2$. Then f is not a product of two polynomials of positive degree in $A[X]$.

P.184, proof of second example [=]: Why is it true that ''It will suffice to prove that the polynomial $f(X+1)$ is irreducible over $\mathbf{Q}$''? Roughly speaking, because a factorization $f(X) = g(X)\,h(X)$ would imply a factorization $f(X+1) = g(X+1)\,h(X+1)$, so if no factorization of the latter sort exists, no factorization of the former sort does either.

The elegant way to say this is that the homomorphism $\mathbf{Q}[X] \to \mathbf{Q}[X]$ taking X to $X+1$, an instance of the evaluation homomorphisms defined on p.98, has an inverse, the evaluation homomorphism taking X to $X-1$. Hence that map is an automorphism of $\mathbf{Q}[X]$, hence preserves irreducibility.

P.184, last display [=]: The symbol ''...'' in the numerator of the right-hand term may be deceptive if you don't stop and think about. It does not mean that the omitted coefficients are all p; rather, these coefficients are given by the Binomial Theorem.

P.186, top paragraph [=]: Hint for proving the proposition: Gauss's Lemma.

Re §IV.4. Hilbert's Theorem.

P.186, first paragraph [>]: For basic results on Noetherian rings, see the discussion of the topic in this Companion under §II.5 above. The result cited by Lang from Chapter X in the proof of Theorem IV.4.1 (the equivalence of the three conditions on p.413) is a generalization to modules of the first Lemma of those pages; but that Lemma is all that is needed in this case.

Theorem IV.4.1, here called ''Hilbert's Theorem'', is commonly known as the **Hilbert Basis Theorem**. In early usage, ''basis'' meant ''generating set'', and the theorem is equivalent to the statement that if every ideal of A has a finite generating set, so does every ideal of $A[X]$. Nowadays, in linear algebra ''basis'' only means a *linearly independent* generating set, but the old usage is fossilized in the common

name for this result.

P.187, last sentence of the proof of Theorem IV.4.1 [=]: Change this to ''By induction, the polynomial $f - c_1 f_{d1} - \ldots$ lies in the ideal generated by the f_{ij}, from which we conclude that f does as well.''

P.187, Corollary IV.4.2 [=]: Note that while the X of the preceding theorem was an indeterminate, the $x_1, \ldots, x_m$ of the Corollary are arbitrary ring elements.

Re §IV.5. Partial fractions. We skip this section in 250A. However, here is one erratum:

P.187, bottom line [=]: After ''$j(p) = 0$'' add ''*and* $\alpha_p = 0$''.

Re §IV.6. Symmetric polynomials.

P.190, beginning of §IV.6 [<]: The results of §§IV.2-3 were aimed at helping us study the possible roots of a polynomial in terms of its coefficients. But one can often get additional information about a problem by turning it backward. Thus, in this section, we look at a polynomial which *is* a product of linear factors, perhaps in some extension of A, and we examine how its coefficients depend on its roots. In general, the roots might be any elements of any commutative A-algebra; but if we write down a polynomial f whose roots are distinct *indeterminates*, i.e., generators of a polynomial ring, $t_1, \ldots, t_n \in A[t_1, \ldots, t_n]$, then from the coefficients of the polynomial

$$F(X) \;=\; (X - t_1) \ldots (X - t_n) \;\in\; A[t_1, \ldots, t_n][X],$$

we can get the coefficients of any polynomial of the form $(X - a_1) \ldots (X - a_n)$ by substituting the a_i for the t_i. (Here ''substituting'' means applying the unique homomorphism of A-algebras that takes t_i to a_i. If the a_i lie in A, this will be a homomorphism $A[t_1, \ldots, t_n] \to A$; if, more generally, they lie in a commutative A-algebra B, it will be a homomorphism $A[t_1, \ldots, t_n] \to B$.) Results about the polynomial $(X - t_1) \ldots (X - t_n)$ will thus give us information about any polynomial that can be written as a product of linear factors $X - a_i$.

P.190, end of first paragraph [>]: The arguments of this section do not require us to know what the s_i look like, and Lang only shows us s_1 and s_n, but the general case is easy to describe: s_i is the sum of the $\binom{n}{i}$ monomials which are products of i distinct factors taken from $\{t_1, \ldots, t_n\}$. E.g., when $n = 4$,

$$s_2(t_1, t_2, t_3, t_4) \;=\; t_1 t_2 + t_1 t_3 + t_1 t_4 + t_2 t_3 + t_2 t_4 + t_3 t_4.$$

P.191, lines 3-5 [>]: The idea of the ''weight'' of a polynomial is ''the degree one would expect the polynomial gotten by substituting for each X_i the symmetric function s_i (or any other polynomial of degree i) to have''. Thus, this is an ad hoc definition, specific to this context. In other contexts, one may use ''weight'' functions defined by associating a different family of weights to a family of indeterminates.

Since the degree of the zero polynomial was defined to be $-\infty$, we shall also understand this to be the weight of the zero polynomial (though Lang does not say this).

One sometimes wishes to speak of a polynomial all of whose monomials have the same weight with respect to some weight function. The formal term for such a polynomial is ''isobaric'', though one more often hears ''homogeneous with respect to the given weight function''.

P.191, 4th line of *Proof* [=]: To ''substitute $t_n = 0$'' here means to apply the homomorphism $A[t_1, \ldots, t_n] \to A[t_1, \ldots, t_{n-1}]$ taking t_n to 0 and all other t_i to themselves.

P.192, first line [=]: After ''By induction'' add ''on d''. (There are two inductions going on in this proof, one on n, and one on d.)

P.192, line 6 [=]: Where Lang writes ''each term on the right has weight $\leq d$'', he means that $g_1(X_1, \ldots, X_{n-1})$ and $X_n g_2(X_1, \ldots, X_n)$ each have weight $\leq d$. Defining $g(X_1, \ldots, X_n) = g_1(X_1, \ldots, X_{n-1}) + X_n g_2(X_1, \ldots, X_n)$, we conclude that g has weight $\leq d$, giving the conclusion of the theorem as Lang states it. However, note that if g had weight strictly less than d, then on substituting s_i for each X_i, we would get a polynomial of degree strictly less than d in the t_i's, contradicting our assumption on f. Hence the weight of g must be exactly d. (The reason we could only say about each of g_1 and $X_n g_2$ that it had weight $\leq d$ is that one or the other of them might have smaller weight;

e.g., you might work out what happens when $f(t) = s_1$, respectively $f(t) = s_n$.)

P.192, italicized statement [=]: This is equivalent to saying that the polynomial g of the theorem just proved is unique. (So you could add the word ''unique'' to the theorem, and consider this the last part of the proof.)

In the proof of this statement, Lang implicitly uses induction on n again.

P.192, end of above proof (up to **Example**) [>].

Alternative proof of Theorem IV.6.1.

The preceding proof is instructive: on substituting $t_n = 0$, one loses some information about one's symmetric polynomial, but reduces to a case in which one can inductively assume the desired result. Thus, one can represent the resulting polynomial in terms of the elementary symmetric functions in $n-1$ variables. Replacing these by the corresponding symmetric polynomials in n variables, we get polynomial which, on the one hand, is expressible in terms of the s_i's, and on the other, resembles our original polynomial to a certain extent. This turns out to mean that the difference between the original polynomial and the new one is s_n times a symmetric polynomial of lower degree, and an induction on degree then allows us to get the desired result.

Here is a different (and also instructive) way to prove the same theorem. (In stating it, I will use ''monomial'' to mean what Lang calls a ''primitive monomial'', i.e., a member of the monoid generated by the elements in question, without coefficient from the base-ring.)

The symmetric group S_n permutes the monomials in $t_1, \dots, t_n$, and the symmetric polynomials in these indeterminates are those polynomials f such that for each *orbit* of this action, the coefficients in f of all monomials in the orbit are the same. Let us select one representative from each such orbit. To do this, we consider the monomials of each degree to be ordered *lexicographically*, i.e., we write $t_1^{d_1} \dots t_n^{d_n} > t_1^{e_1} \dots t_n^{e_n}$ if and only if the least i for which $d_i \neq e_i$ has $d_i > e_i$. (In other words, in comparing monomials of the same degree, we compare first the exponents of t_1, then, if these are equal, we compare the exponents of t_2, etc..) We can see that in each orbit of monomials under the action of S_n, the *greatest* monomial

(c27) $$t_1^{d_1} \dots t_n^{d_n}$$

under this ordering will be the unique member of the orbit which satisfies

(c28) $$d_1 \geq d_2 \geq \dots \geq d_n.$$

Moreover, it is not hard to verify that for each monomial (c27) satisfying (c28), there is a unique monomial in $s_1, \dots, s_n$ having (c27) as its greatest term under our ordering; namely

(c29) $$s_1^{d_1 - d_2} s_2^{d_2 - d_3} \dots s_n^{d_n}.$$

Now suppose f is any symmetric polynomial in $t_1, \dots, t_n$. Let us look at its highest-degree monomials, and among these, find the monomial (c27) that is greatest under our ordering. Since f is symmetric, this monomial will satisfy (c28), hence if we subtract (c29) times the coefficient of this monomial, we get a symmetric polynomial whose greatest monomial of highest degree is either of smaller degree than that of f, or of the same degree, but smaller under our ordering. Repeating this process, we must sooner or later run out of monomials of the original highest degree (since there are only finitely many monomials of that degree); we then eliminate monomials of the next lower degree the same way, and so on. This process must stop when the subtraction at some step yields remainder 0. When this happens, we will have expressed f as a polynomial in $s_1, \dots, s_n$.

The same method also yields an easy proof of the italicized assertion on p.192 following the proof of the theorem (which should really be part of the statement of the theorem): it follows from the uniqueness statement in the line above (c29) that distinct monomials in $s_1, \dots, s_n$ involve distinct highest monomials in $t_1, \dots, t_n$ (under our ordering). It follows that a nontrivial linear combination of monomials in

$s_1, \dots, s_n$ can never equal zero; for if we look at the highest monomial in $t_1, \dots, t_n$ occurring among their expansions, this can only occur in one of them, hence will not be canceled when we sum these terms.

Having seen Theorem IV.6.1 proved in two ways, you should now convince yourself that a little bit more is true than what that theorem says; namely that if $f(t) \in A[t_1, \dots, t_n]$ is symmetric and *homogeneous* of degree d (cf. first of my comments on p.191 above), then it can be written as $g(s_1, \dots, s_n)$ where g is homogeneous (or in purist language, isobaric) of degree d with respect to our weight function. This will be used on p.193. One way to get this result is to go back over either of the above proofs and verify that they work with ''homogeneous'' inserted everywhere. Another is to use the result as stated to get the existence of a not-necessarily-homogeneous polynomial with the stated property, then group together its monomials homogeneous of each weight getting an expression $g = g_d + g_{d-1} + \dots + g_0$, note that since each elementary symmetric polynomial s_i is homogeneous of degree i in $t_1, \dots, t_n$, the substitution of the s's for the X's gives an expression for f in terms of homogeneous summands $g_i(s_1, \dots, s_n)$; but now since f was assumed homogeneous of degree d, the summands of degrees other than d must be 0, and we must have $g = g_d$, as claimed.

P.193, first paragraph [>]: Lang uses the displayed formula to define a polynomial D_f, the unique polynomial in n indeterminates such that, when one substitutes for these indeterminates the respective elementary symmetric functions in $t_1, \dots, t_n$, one gets the product of the squares of the pairwise differences of the t_i. Now if we map $A[t_1, \dots, t_n]$ into an arbitrary A-algebra B by sending $t_1, \dots, t_n$ to any elements $a_1, \dots, a_n$, the above characterization of D_f yields the following statement: Given any polynomial of degree n over an A-algebra, which is a product of linear terms, $(X - a_1) \dots (X - a_n)$, if one substitutes for the indeterminates in D_f the coefficients of this polynomial, one gets the product of the squares of the differences of the roots a_i.

It is this property that Lang will use for the rest of the section. In fact, whenever he specifies a polynomial f, he will then use the symbol ''D'' for the element gotten by substituting the coefficients of that polynomial for the indeterminates in the above polynomial D_f. Thus, under **Quadratic case** he will take $n=2$, look at the polynomial $f(X) = X^2 + bX + c$, and understand D to mean $D(b, c)$, while under **Cubic case** he will take $n=3$, look at the polynomial $f(X) = X^3 + aX + b$ (note the absence of an X^2 term), and understand D to mean $D(0, a, b)$.

Note also that $\prod_{i<j} (t_i - t_j)^2$, in addition to being symmetric, is homogeneous of degree $n(n-1)$; hence by our earlier observations, $D(s_1, \dots, s_n)$ will be homogeneous of weight $n(n-1)$. This property will be retained if we replace one or more of the s_i by 0; in particular, $D(0, a, b)$ is homogeneous of weight 6. Lang will use this fact, without clearly stating it, a couple of paragraphs from now.

P.193, **Cubic case** [=]: Note that by taking the coefficient of X^2 to be 0, as noted above, Lang is essentially taking $t_3 = -t_1 - t_2$. This is equivalent to looking at the images of s_1, s_2, s_3, D, etc. under the map $A[t_1, t_2, t_3] \to A[t_1, t_2]$ taking t_3 to $-t_1 - t_2$; and indeed, Lang here speaks of D as a polynomial in t_1 and t_2. (He will indicate in the top paragraph of the next page how, by a substitution, one can reduce the problem of finding the discriminant of an arbitrary cubic to the case of a cubic with no X^2 term.)

Note that in the context of the preceding pages, each symmetric polynomial s_i is homogeneous of degree i in $t_1, \dots, t_n$, δ is homogeneous of degree $n(n-1)/2$ in those indeterminates, and hence D is homogeneous of degree $n(n-1)$. When $n=3$ and we map $k[t_1, t_2, t_3]$ onto $k[t_1, t_2]$ by sending t_3 to $-t_1 - t_2$ as described above, the latter element is homogeneous of degree 1 in t_1 and t_2, so homogeneous elements remain homogeneous of the same degrees. This allows Lang to use homogeneity arguments in determining the form of D.

Three lines after middle display on this page, for both occurrences of the word ''special'' read ''particular''.

In the second line after the next display, ''of the product'' should be ''as the product''.

Re §§IV.7-9. This is material which there is no time to cover in 250A; but I give below some clarification of material on pp.196-197 brought to my attention by David Goldberg.

P.196, 3rd line after 2nd display [=]: Where Lang speaks of ''large'' primes, I believe the point is that the product $N_0(abc)$ of the primes in question must be large, hence they can't just be among the first few primes; so the largest of them must be at least moderately large. (But not necessarily anywhere near abc.)

P.197, 2nd line after first display [=]: For ''Again the proof is immediate from'', read ''We shall show below that the above conjecture is implied by''.

P.197, 3rd line before display (1) [=]: $mv >$ should be $mn >$.

P.197, display (1) [=]: Sketch of the derivation of these formulas: In the *abc* conjecture we may bound $\max(|a|, |b|, |c|)$ below by $|a|$, and $N_0(abc)$ above by $N_0(a)N_0(b)N_0(c)$. If we apply the resulting inequality to the equation preceding (1), dropping the factors A and B because they are fixed, bounding $N_0(u^m)$ and $N_0(v^n)$ by u and v respectively, and writing the result in ''<<'' notation, we get $|u|^m << (|u||v|N_0(k))^{1+\varepsilon}$, hence $|u|^{m-1-\varepsilon}|v|^{-1-\varepsilon} << N_0(k)^{1+\varepsilon}$. We get a similar inequality with the roles of u and v reversed, and taking a ''multiplicative linear combination'' of these which eliminates $|v|$, and solving for $|u|$, we get $|u| << N_0(k)^{(1+\varepsilon)n/(mn-(m+n)(1+\varepsilon))}$. The exponent is not the same as in Lang's (1), but given any $\varepsilon > 0$, we can find an $\varepsilon' > 0$ (independent of m and n so long as $mn/(m+n)$ is uniformly bounded above 1; and indeed the relation $mn > m+n$ leads to a bound $6/5$) such that $(1+\varepsilon')n/(mn-(m+n)(1+\varepsilon')) \leq (1+\varepsilon)n/(mn-(m+n))$. Hence, applying the above computation with that ε', we get the first formula of (1). The second formula follows by symmetry.

P.197, sentence after next-to-last display [=]: For this read, ''The Hall conjecture is essentially the case of the first formula of (1) with $m=3$, $n=2$, and $N_0(k)$ bounded by $|k| = |u^3 - v^2|$.''

Chapter V. Algebraic Extensions.

Re §V.1. Finite and algebraic extensions.

P.223, first paragraph of §V.1 [>]: In addition to referring to a field E as an *extension* of a field F, Lang will frequently refer to the pair $F \subseteq E$ as ''an extension''.

Note that ''E is a finite extension of F'' does *not* mean E is a finite field.

P.224, top paragraph [~]: Here $F[\alpha]$ means the ring generated over F by α; cf. the notation $A[S]$ introduced on p.90.

Here and subsequently Lang will speak of polynomials ''with leading coefficient 1''. There is a standard term for such polynomials, which he introduced on p.175, but unfortunately rarely uses: *monic polynomials*.

What Lang calls ''the irreducible polynomial of α over F'' and writes $\mathrm{Irr}(\alpha, F, X)$ is generally called the *minimal polynomial* of α over F. Note the three arguments of ''Irr''. The significance of α is clear; the argument F is needed because a given element α will have different minimal polynomials over different fields (e.g., i has minimal polynomial X^2+1 over the real numbers, but $X-i$ over the complexes). Finally, if we are dealing with more than one indeterminate, then simply saying ''the minimal polynomial of i over the reals'' would not show whether we meant X^2+1 or Y^2+1, so the third argument is also necessary.

The degree of $\mathrm{Irr}(\alpha, F, X)$ is called the **degree** of α over F. Likewise, the number $[E:F]$ that Lang associates with the extension E of F later on this page is called the **degree** of the extension. (The index refers to this page for the term ''degree'', though the word does not appear on the page.)

P.225, Corollary V.1.3 [=]: This should begin with the same sentence as the preceding proposition.

P.225, proof of Proposition V.1.4 [<]: The first statement of the proposition is equivalent to saying that $k[\alpha]$ is a field. There is an easier proof than Lang gives, using the first display on the preceding page. For since $(p(X))$ is an ideal of a principal ideal domain generated by an irreducible element, it is a *maximal* ideal, so $k[X]/(p(X))$ will be a field.

P.226, second paragraph [>]: Note that where Lang defines the compositum EF, it is not asserted that this consists of products of elements of E and elements of F, or even of sums of such products. (What form *will* its elements have?) In the last line of that paragraph, for ''we cannot define the compositum'' read ''the compositum is not defined''; i.e., the above definition is not applicable.

You might find it interesting to prove that given subfields E, F of a field L, if we write $k = E \cap F$, then if either E or F is algebraic over k, their compositum EF will be generated by E and F as a ring, i.e., will consist of sums of products of elements of E and elements of F. On the other hand, for an example of a compositum *not* having this form, start with a field k, let $L = k(s,t)$ where s and t are algebraically independent over k, and let $E = k(s)$, $F = k(t)$. You will find that in the ring generated by E and F, $s - t$ is not invertible; hence the compositum EF is bigger than that ring.

(The word ''compositum'' is sometimes used in a slightly different way, though not in Lang. Given extensions E and F of a field k, ''*a* compositum of E and F'' means any extension L of k given with homomorphisms, i.e., embeddings, of E and F into it over k, such that L is generated by the images of those embeddings. Under that definition, one can prove that any two extensions of k *do* have a compositum, though it is not, in general, unique, even up to isomorphism.)

P.226, middle paragraph [>]: Compare this with the notation introduced in the third paragraph of p.110, where $K(X_1, \dots, X_n)$ is defined as the field of fractions of the polynomial ring $K[X_1, \dots, X_n]$. The convention of which these are both cases is that when we list some elements in parentheses after the name of a field, we are naming a field extension generated by those elements. Whether these elements are independent indeterminates, as on p.110, or elements of a given larger field, as on this page, must be

deduced from context (sometimes from the symbols used). If there is danger of confusion, the writer should say explicitly what is meant.

In the next paragraph Lang will define a field E to be *finitely generated* over a subfield k if and only if it can be written $k(\alpha_1, \dots, \alpha_n)$ for some finite family of elements $\alpha_1, \dots, \alpha_n$ of E. However, such an extension will not necessarily be finitely generated over k as a *ring*. This is because an extension $k(\alpha_1, \dots, \alpha_n)$ may or may not coincide with its subring $k[\alpha_1, \dots, \alpha_n]$.

P.227, last paragraph [~]: Lang's term ''distinguished class'' is essentially a mnemonic: Various classes of field extensions share certain properties, so if we call a class of extensions having these properties ''distinguished'', then instead of having to memorize this list of properties for each such class, we simply have to recall that we have proved that the class is distinguished. Note, however, that this definition differs from the commonest mathematical use of the word ''distinguished'', where, for instance, if X is a set and x an element of X, then the pair (X, x) may be referred to as ''a set with a distinguished element''.

(I would suggest replacing Lang's phrase ''$\mathcal{C}$ is a distinguished class'' with the phrase ''$\mathcal{C}$ has the three standard properties''.)

P.228, third paragraph [~]: Here ''E/F'' means ''E, viewed as an extension of F''. This allows us to abbreviate statements like ''E is algebraic over F'' and ''E is finite over F'' to ''E/F is algebraic'' and ''E/F is finite''. This notation is commonly used, though its similarity to the notation for factor-groups, factor-rings etc. is unfortunate. (Stewart's little book *Galois Theory* introduces instead the symbol $E:F$, which does not have this disadvantage.)

P.229, end of §V.1 [>]: Here are some further notes on the material in this section.

Hints of what is to come.

Algebraic extensions of fields are curious in that their structure is neither ''very tight'' nor ''very loose''. If we map an algebraic extension field E of a field k into another extension F of k, then for each $\alpha \in E$, the element of F to which it is mapped must be a root of the minimal polynomial of α over k. A polynomial of degree n has at most n roots in a field; thus, though there may be more than one element of F to which α can be sent, there cannot be more than $\deg_k(\alpha)$. This ''semi-rigid'' structure on homomorphisms among algebraic extension fields can be best investigated by examining the group of *automorphisms* of a single field extension F of k. We can see from the above that if F is finite over k, then $\mathrm{Aut}_k(F)$ is finite, and we shall find that its structure is intimately connected with the structure of the extension. This is the subject of Galois Theory.

In the above paragraph we used the fact that an irreducible polynomial of degree n over k has *at most* n roots in an extension field of k. Such a polynomial usually has a full complement of n distinct roots in a big enough extension of k. Usually – but not always! The situation where it does not is called ''inseparability'', and will be studied in §V.6. Let us note one example here. Suppose k_0 is a field of finite characteristic p, let t be an indeterminate over k_0, let $L = k_0(t)$, and define $k = k_0(t^p) \subseteq L$. Then as an extension of k, L is generated by the element t, which is a root of the polynomial $X^p - t^p \in k[X]$. In $L[X]$, this polynomial splits into linear factors, $X^p - t^p = (X - t)^p$; but we see from this factorization that it has only one root, t. Galois theory is limited to the study of extensions in which this phenomenon does not occur. These are called *separable* extensions, and introduced in §V.4.

On fields, and universal constructions.

We encountered the concepts of ''product'' and ''coproduct'' in the study of groups and rings, and it is natural to ask whether such constructions exist for fields. In general, the answer is no. Given fields E and F, if we form their product as rings, this will have zero-divisors $(0,1)\cdot(1,0) = 0$, and so will not be a field; and, in fact, there is in general no field with the required universal property. Likewise, if we try to construct a coproduct by looking at what elements we can name in a field with homomorphisms of E and F into it, we find that we can name all the *ring-theoretic* expressions in such elements, but to get a field,

we must also have inverses of all *nonzero* elements. However, a given ring-theoretic expression in elements of E and F might have the value zero in one field L_1 containing copies of E and F, but be nonzero in another such field L_2. Hence if we introduce an inverse to this expression, we lose universality (we can't map to L_1), while if we don't, we don't get inverses of all nonzero elements, and so don't have a field.

Why this difference between fields and other classes of algebraic objects? The other classes we have seen can each be defined as consisting of all sets given with certain everywhere-defined operations, subject to certain fixed identities. (To put the definitions in this form sometimes requires a reformulation. For instance, the condition that every element of a group have an inverse is not an identity. But if we define a group in terms of *three* operations, composition (a binary operation), inverse (a unary operation) and neutral element (a ''zeroary'' operation), we can then state *all* the group axioms as identities.) Such a class of objects is known in universal algebra as a *variety of algebras*, and one can prove that in any variety, one has free objects, products, coproducts, etc.. The class of fields fails to be a variety in that the multiplicative inverse operation is not everywhere defined – it is defined only on nonzero elements. Another important class of algebraic objects that fails to be a variety is the class of *integral domains*, since the condition ''$xy = 0 \Rightarrow x = 0$ or $y = 0$'' is not an identity.

It is interesting to note, however, that if E and F are extensions of a field k, and we form their coproduct $E \amalg F$ as commutative k-algebras, then all field extensions of k generated by an image of E and an image of F can be constructed, up to isomorphism, as the *fields of fractions* of the *factor-rings* $E \amalg F/P$, as P ranges over all prime ideals of $E \amalg F$.

Quaternions.

Though the present chapter concerns fields, I would like to briefly introduce you to one important example of a noncommutative division ring: the ring of quaternions.

This is an algebra over the real numbers, generally denoted $\mathbf{H}$ (for its discoverer, William Rowan Hamilton), which is 4-dimensional as an $\mathbf{R}$-vector-space, with basis denoted $\{1, i, j, k\}$, and with multiplication determined by the conditions that 1 be the multiplicative neutral element, and that the other basis elements satisfy

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j. \tag{c30}$$

Thus, the general element of $\mathbf{H}$ has the form $a + bi + cj + dk$ $(a, b, c, d \in \mathbf{R})$, and a product of two such elements is calculated by applying bilinearity over $\mathbf{R}$, and evaluating each of the resulting summands using (c30). To verify that this multiplication is associative, it suffices to check the associative law for each of the $4 \cdot 4 \cdot 4 = 64$ products in which each term is taken from $\{1, i, j, k\}$. Actually, it is easy to see that associativity will be automatic whenever one of the factors is 1; so we are reduced to the 27 cases where each term is taken from $\{i, j, k\}$; and in fact, since (c30) is invariant under cyclic permutations of $\{i, j, k\}$, the problem reduces to checking the nine cases in which the first factor is i; a not unreasonable amount of work.

Once we have shown that $\mathbf{H}$ is an associative algebra, how do we prove that every element is invertible? By a trick very much like that used for the complex numbers. Given a quaternion $\alpha = a + bi + cj + dk$, define $\bar{\alpha} = a - bi - cj - dk$. Then it is easy to check that $\alpha\bar{\alpha} = a^2 + b^2 + c^2 + d^2$. (Do you see why the ''mixed terms'', e.g., those with coefficient ab, or bc, all cancel?) Since this real number $\alpha\bar{\alpha}$ is invertible, an inverse to α is given by $\bar{\alpha}/\alpha\bar{\alpha}$.

Some miscellaneous observations on the quaternions: It follows from (c30) that the eight elements ± 1, $\pm i$, $\pm j$, $\pm k \in \mathbf{H}$ form a group; this is the group Lang calls the ''quaternion group'' on p.9, last paragraph (though he writes m in place of -1, since ''-1'' is not meaningful in an abstract group). Note also that the second line of (c30) agrees with the ''cross product'' operation on $\mathbf{R}^3$, although, of course, the first line does not. The division ring $\mathbf{H}$ clearly contains three copies of the complex numbers, $\mathbf{R}(i)$, $\mathbf{R}(j)$

and $\mathbf{R}(k)$. In fact, it contains uncountably many such copies, since it is not hard to show that any quaternion $\alpha = bi+cj+dk$ such that $b^2+c^2+d^2 = 1$ satisfies $\alpha^2 = -1$, and generates a commutative subfield of $\mathbf{H}$ isomorphic to $\mathbf{C}$.

Clearly $\{\, a+bi+cj+dk \mid a, b, c, d \in \mathbf{Z} \,\}$ is a subring of $\mathbf{H}$; but in fact, it is easy to verify that the larger subset consisting of all elements $(a+bi+cj+dk)/2$ such that a, b, c, d are integers all congruent to one another mod 2 also forms a subring, and this turns out to have nicer arithmetical properties. It is therefore the latter ring that is called the ring of *integer quaternions*.

There are other important examples of noncommutative division rings than $\mathbf{H}$, but none as easily described or as well-known, and it was in fact the first to be discovered.

Re §V.2. Algebraic closure.

P.229, beginning of §V.2 [<]: Be prepared for some messiness here. We will be dealing with situations that would look neater if we identified isomorphic objects, but instead, we will explicitly name isomorphisms between them.

Given an isomorphism between two objects, isn't it always safe to make an identification, so as to regard the two objects as the same, and the isomorphism as the identity? It frequently is; but a big exception is when the objects are already the same, or lie within a common structure being considered, since then making such an identification would amount to considering distinct elements of that structure as the same. And, in fact, we will soon be using the results of this section in the case where the isomorphisms referred to are automorphisms (or partial automorphisms) of one field.

Nevertheless, until such assumptions are brought in, in understanding results that Lang states for a given embedding σ of one field in another, it can be useful to first picture the case where the embedding is an inclusion.

The need to avoid treating isomorphic objects as the same leads Lang to make such distinctions in this section even in cases when the two objects will not in fact be part of a common object; for instance in not identifying a field k with its canonical image in a field $k[X]/I$ (I an ideal of $k[X]/I$) at the bottom of p.230. However, thinking how to handle such distinctions rigorously is good practice for us.

P.229, next-to-last paragraph [>]: Note that Lang not only writes a^σ for $\sigma(a)$, but also f^σ for the polynomial gotten by applying σ to every coefficient of f, while leaving the indeterminate X unchanged, so that one is still free to substitute an arbitrary element for that indeterminate. (In the notation introduced on p.99, next-to-last display, this polynomial would be written (σf). Another way of formulating the final display on this page would be $(\sigma f)(\tau\alpha) = (\tau f)(\tau\alpha) = \tau(f(\alpha)) = \tau(0) = 0$.)

P.230, proof of Lemma V.2.2 [~]: This can be done without any computation. We know that E_1E_2 can be characterized as the least subfield of E containing E_1 and E_2, and that $\sigma(E_1)\sigma(E_2)$ is likewise the least subfield of L containing $\sigma(E_1)$ and $\sigma(E_2)$. From this, and the fact that σ gives an isomorphism of E with $\sigma(E) \subseteq L$, it immediately follows that it takes E_1E_2 to $\sigma(E_1)\sigma(E_2)$.

P.230, 5 lines from bottom [~]: Here Lang pauses to prove something he should have noted (and almost did) on p.93: That any homomorphism from a field into a nonzero ring has kernel 0.

Setting aside that distraction, note the point of this paragraph in Lang: he shows that, given a polynomial f over a field k, one can *construct* an extension of k in which f has a root. You have probably seen this argument in an undergraduate algebra course; if not, take careful note of the method. There is a minor technical complication, mentioned earlier, that the field one gets does not contain k itself, but an isomorphic copy σk thereof; this will be patched up in Proposition V.2.3, but meanwhile, it causes the result proved here to take the less transparent form "$p^\sigma(\xi) = 0$" instead of "$p(\xi) = 0$". Recall that $p^\sigma(X)$ is the polynomial obtained by applying σ (equivalently, the restriction of σ to k) to all the coefficients of p, but not to the indeterminate X.

It may seem confusing to have a polynomial in which the coefficients are elements of $k[X]/(p(X))$, and yet in which X appears as the indeterminate. If so, you might find the reasoning easier to follow if you give the indeterminate in p^σ a different name, writing something like $p^\sigma(Y)$. (Lang doesn't actually

write ''$p^\sigma(X)$'' in the proof, but he writes p^σ; and in thinking about the argument, one should be able to refer to the indeterminate in that polynomial.)

P.231, end of proof of Proposition V.2.3 [>]: It would actually have been better to have proved a general lemma that if $\sigma\colon k \to F$ is an embedding of fields, then there is an extension E of k such that σ extends to an isomorphism $E \cong F$, since this is what the proof really shows. Lang will use this fact again on the next page, where he speaks of ''using the same type of set-theoretic argument as in Proposition 2.3''.

P.231, proof of Theorem V.2.5 [>].

Remarks on the construction of algebraically closed field extensions

To construct an algebraically closed extension field of k, Lang first adjoins a root of every polynomial over k, then does the same for every polynomial over the resulting field, and so on, iterating this countably many times, and finally takes the union. This is a useful general technique to be aware of. Actually, in *this* case, the successive iterations are not needed; one can prove that an algebraic extension of k in which every polynomial over k has a root is already algebraically closed. (This is Exercise VI:L28, p.325.) But at this point in the development, it is easier to invoke the countable chain of extensions than to prove that one iteration suffices!

One might ask whether one can't prove the existence of an algebraically closed extension of any field k by Zorn's Lemma. Clearly the union of any chain of algebraic extensions of k is again an algebraic extension. Doesn't Zorn's Lemma therefore give a maximal such extension? The trouble with this approach is that the algebraic extensions of k do not form a *set*. (Zorn's Lemma was only proved for partially ordered *sets*; and, indeed, if the restriction to sets could be removed, one could use the resulting ''lemma'' to prove such statements as that there was a largest set, which is absurd.)

But with some additional work, one can indeed get a Zorn's Lemma proof based on this idea. Given k, the first step is to get a bound on the cardinalities of algebraic extensions of k. I claim that such a bound is given by the cardinality of the polynomial ring $k[X]$. Indeed, any algebraic extension E of k can be mapped set-theoretically to $k[X]$ by sending each element to its minimal polynomial, and this map is finitely-many-to-one, since each polynomial has only finitely many roots in E. But a set A which can be mapped in a finitely-many-to-one fashion to an infinite set B has cardinality no greater than that of B, giving the asserted bound. (Lang will sketch approximately the same argument very briefly on p.235, though not for the purpose of giving an alternative construction of algebraically closed fields.) In view of this result, we can find a set S whose cardinality is *greater* than that of any algebraic extension of k. Without loss of generality, let us assume that S contains k. We may now let P be the set consisting of all field extensions of k whose underlying sets are contained in S, and partially order P by the relation ''is an extension of''. It is easy to verify that this set is inductive; thus P has a maximal element, K.

It remains to prove that K is in fact algebraically closed. Assuming the contrary, let L be a proper algebraic extension of K. This will still have cardinality less than that of S. Moreover, since K has cardinality less than that of S, the cardinality of $S - K$ is the same as that of S; in particular, we can map the set $L - K$ bijectively to some subset $U \subseteq S - K$. Using the identity map on K and the above bijection on $L - K$, we get a bijection between L and the subset $K \cup U$ of S, which is the identity on K. We can use this bijection to transport the field structure of L onto $K \cup U$. Since L is a proper algebraic extension of K, $K \cup U$ with this field structure will be a proper algebraic extension of K whose underlying set lies in S. But this contradicts the maximality of K, so we conclude that K is indeed algebraically closed.

(In setting up the above proof, we might have thought it would be enough to take S to have cardinality greater than *or equal to* that of any algebraic extension of k; e.g., denumerable if k was so. You should look carefully at the above proof, and see why this would not have worked.)

I will make some remarks at the end of this section on the idea behind Lang's proof of the existence of an algebraically closed extension of k, in particular, the first part, where it is shown that one can

simultaneously adjoin roots to all polynomials of positive degree over k.

P.233, Proposition V.2.7 [~]: In the preceding paragraphs, Lang has proved more than he records in the proposition, and in fact, he will use the stronger result repeatedly, so you should note it: *If $E = k(\alpha)$, where α is algebraic over k, with minimal polynomial p, and $\sigma\colon k \to L$ is an embedding of k in a field L, then the extensions of σ to E correspond bijectively to the distinct roots of p^σ in L; namely, for each such root β, there is a unique extension of σ carrying α to β.* (In particular, L does not have to be algebraically closed.)

In many applications of this result, L will also be an extension of k, and σ the inclusion map. Note that in that case, the conclusion simplifies to: *Homomorphisms $E \to L$ over k correspond bijectively to roots of p in L; namely, for each such root β, there is a unique such homomorphism carrying α to β.*

P.234, Corollary V.2.9 [=]: In the proof (though not in the statement) of the corollary, the ''identity mapping on k'' should be ''the inclusion $k \to E'$''.

P.235, second and third paragraphs [<]: These are confusingly written, and can be skipped.

P.235, end of §V.2 [>]:

''*Compactness*'' *results* (sketch, optional).

Let us look again at Lang's proof that a field k admits an extension in which every polynomial $f \in k[X]$ of positive degree has a root. He begins by showing that for every *finite* set of such polynomials f, one can get an extension of k in which these all have roots. He then adjoins to k an infinite family of indeterminates X_f, one for each f, and forms the ideal I generated by the elements $f(X_f)$. To complete the proof by dividing out by a maximal ideal containing I, we need to know that I is a *proper* ideal of the polynomial ring.

To get this, Lang observes that if it contained 1, then the subideal generated by some finite family $f_1(X_{f_1}), \dots, f_n(X_{f_n})$ would also. But one can construct a homomorphism from our polynomial ring into a field in which $f_1, \dots, f_n$ all have roots, which has $f_1(X_{f_1}), \dots, f_n(X_{f_n})$ in its kernel; hence the ideal generated by these elements does not contain 1, giving the result needed for the proof.

In general, an argument showing that if a set S of conditions expressible in an appropriate language has the property that every *finite* subset of S can be satisfied simultaneously, then all of S can be satisfied simultaneously, is called by logicians a *compactness* theorem; and they have compactness theorems for various sorts of languages. To see the reason for the term, let L be a language, and M the class of all ''models'' of L (structures in which we can interpret the statements of the language. E.g., in our case, the language would be one with symbols for addition, multiplication, and the elements of k, and with variable-symbols, quantifiers, etc., sufficient to express the condition that a structure forms *a field containing k*, and such that we can write, for any $a_0, \dots, a_{n-1} \in k$, the condition

$$(\exists\, x)\ \ x^n + a_{n-1}x^{n-1} + \dots + a_0 \ = \ 0.$$

The class M of models would consist of all sets given with operations and elements to be interpreted as the addition, multiplication, and the scalars from k.) Let us topologize M with the least topology such that for each proposition P in the language L, the set of models satisfying P is a closed set. Then the statement that any set S of propositions all of whose finite subsets admit models itself admits a model is equivalent to saying that M is *compact* under this topology! (Well, almost. The trouble is that, as we have described it, M is not a *set*. This can be gotten around in various ways, e.g., by limiting M to consist of models whose underlying set is a subset of a fixed set of large enough cardinality. In any case, logicians generally do not think of the topological background when they use the term ''compactness theorem''.)

Since beginning graduate students are not assumed to have seen any general compactness theorems, Lang goes through an ad hoc ''compactness argument'' for this case.

We remark that if a language is *too* powerful, compactness will generally fail. For instance, suppose a language can express the condition (*P*): ''the set X is finite'', as well as the more elementary conditions (*Q1*): ''X has more than 1 element'', (*Q2*): ''X has more than 2 elements'', (*Q3*): ''X has more than 3 elements'', etc.. Then we see that any finite subset of **{**(*P*), (*Q1*), (*Q2*), (*Q3*), ...**}** can be satisfied simultaneously, but not all of them together; so we do not have compactness for this language. Note that if a language allows one to quantify over functions, then (assuming other more elementary properties), one can say ''there exists a function from X to X which is one-to-one but not onto'', i.e., that X is infinite; and negating this, one can say that X is finite. Hence languages for which one expects compactness results should in general not allow quantification over functions. You can verify, similarly, that compactness fails for a ring-theoretic language in which one can say, ''x is transcendental'' along with more elementary statements about elements of a field.

Re §V.3. Splitting fields and normal extensions.

P.236, line 3 [>]: Lang has defined the **splitting field** of a polynomial. Here are some examples. A splitting field for the polynomial $X^2 - 2$ over $\mathbf{Q}$ is $\mathbf{Q}(\sqrt{2})$, since over this field we have the factorization $X^2 - 2 = (X - \sqrt{2})(X + \sqrt{2})$. On the other hand, a splitting field of $X^4 - 2$ cannot be constructed simply by adjoining to $\mathbf{Q}$ a fourth root of 2. For instance, if we let $\alpha \in \mathbf{C}$ denote the positive real fourth root of 2, then over $\mathbf{Q}(\alpha)$, the polynomial $X^4 - 2$ factors as $(X - \alpha)(X + \alpha)(X^2 + \alpha^2)$, with a quadratic factor that is irreducible over the real numbers, hence over its subfield $\mathbf{Q}(\alpha)$; so $\mathbf{Q}(\alpha)$ is not a splitting field of this polynomial. On the other hand, you should verify that a splitting field of $X^4 - 2$ *is* given by $\mathbf{Q}(\alpha, i)$. (Remember to check the condition that the field in question is generated by roots of the polynomial!)

Another example: Let $f(X) = X^5 - 1$. If we write $\zeta_5 = e^{2\pi i/5}$, then the roots of f are 1, ζ_5, $\zeta_5{}^2$, $\zeta_5{}^3$, $\zeta_5{}^4$, and it follows immediately that $\mathbf{Q}(\zeta_5)$ is a splitting field for f.

Digression: The existence of splitting fields (or of algebraic closures, obtained in the preceding section) gives an interesting alternative proof of the italicized statement on p.192, that if elements $t_1, \dots, t_n$ are algebraically independent, then the symmetric polynomials $s_1, \dots, s_n$ in these elements are also algebraically independent, at least in the case where the base-ring A is an integral domain. To see this, let K be the field of fractions of A, let $K(t_1, \dots, t_n)$ be the field of fractions of $A[t_1, \dots, t_n]$, and let $K[t_1, \dots, t_n]$ be the subring thereof generated by the indicated elements.

Because $t_1, \dots, t_n$ are assumed algebraically independent, $K[t_1, \dots, t_n]$ can be identified with the polynomial ring in n indeterminates over K; so if L is any field extension of K and $\alpha_1, \dots, \alpha_n$ any elements of L, we get a substitution homomorphism $K[t_1, \dots, t_n] \to L$ that fixes elements of K and sends each t_i to α_i. This homomorphism will send the symmetric polynomials in the ts to the symmetric polynomials in the αs, that is, to the coefficients of the polynomial $\prod_i (X - \alpha_i)$; so if the symmetric polynomials in the ts satisfied a nontrivial polynomial relation $f(s_1, \dots, s_n) = 0$, the coefficients c_i of every polynomial of the form $\prod (X - \alpha_i)$ for any n-tuple of elements α_i of any extension field of K would also satisfy $f(c_1, \dots, c_n) = 0$.

But by the existence of splitting fields, every monic polynomial over an extension field of K does have the form $\prod (X - \alpha_i)$ in some larger extension field; and since we can take any n-tuple of elements as the coefficients of $X^{n-1}, \dots, X^0$ in a monic polynomial, this would mean that every n-tuple of elements c_i satisfied $f(c_1, \dots, c_n) = 0$. Taking an L containing algebraically independent elements $c_1, \dots, c_n$, we get an immediate contradiction.

P.237, statement and proof of Theorem V.3.3 [=]: A clearer statement of **NOR 1** is:

NOR 1. *For every embedding σ of K in k^{a} over k, one has $K^\sigma = K$.*

The idea of the equivalent conditions **NOR 1** - **NOR 3** is to avoid the deficiency we saw above in the field gotten by adjoining to $\mathbf{Q}$ a single fourth root of 2, namely, that of containing some, but not all roots of an irreducible polynomial.

In the proof of the theorem, the first paragraph actually shows that **NOR 1** $\Rightarrow$ **NOR 3** $\Rightarrow$ **NOR 2**, and the second paragraph shows that **NOR 2** $\Rightarrow$ **NOR 1**, so the third paragraph, showing **NOR 3** $\Rightarrow$ **NOR 1**, is unnecessary.

Re §V.4. Separable extension.

P.239, beginning of §V.4 [<]: Suppose F is a field, E an algebraic extension of F, and L an algebraically closed field containing F. Then it is natural to study the embeddings of E in L over F.

As a slightly more general situation, we might assume that L does not necessarily contain F itself, but an isomorphic copy $\sigma(F)$, where $\sigma\colon F \to L$ is an embedding of fields, in which case we should understand ''embeddings of E in L *over* F'' to mean embeddings that make commuting diagrams with the inclusion of F in E, and the map $\sigma\colon F \to L$. In a sense, this is no more general than the first situation, since we can reduce it to that situation by mentally ''identifying'' elements of F with their images in L, and thus regarding L as an extension of F. But, as I discussed at the beginning of §V.2, we will want to avoid formally making such identifications, so that we will be able to use these results in the case where F *is* a subfield of L but σ is *not* the inclusion map. Nevertheless, in reading the material on this page (up to Theorem V.4.1), you might first think of L and L' as algebraically closed *extensions* of F, and σ and τ as the inclusions, then reread those paragraphs without making that assumption.

The point Lang is leading up to here is that in looking at the embeddings of E over F into an algebraically closed extension L of F, one gets the same number of maps, independent of the L one uses.

P.240, paragraph after Corollary V.4.2 [<]: Skip this paragraph. (Lang will make this all clear two sections later.)

P.240, last two sentences [~]: These refer to ''multiple roots''. This means multiple roots in an *algebraic closure* of k (or in any field in which the polynomial in question splits). These polynomials do not, in general, have roots in k itself. The same applies to later uses of the phrase.

Incidentally, note that for any finite extension E of a field k within k^{a}, we have

$$\#\text{ of automorphisms of } E \text{ over } k \;\leq\; \#\text{ of embeddings of } E \text{ in } k^{\mathrm{a}} \;\leq\; [E\colon k].$$

By the results of the last section, E is *normal* over k if and only if one has equality at the first $\leq$ sign, while by the definition just given, it is *separable* over k if and only if one has equality at the second. So normality plus separability means that it has as many automorphisms as one could hope for. We will see in the next chapter that this pair of conditions defines an important class of extensions.

P.241, proof of Theorem V.4.3 [>]: Some further basic facts that you should be able to verify at this point, but which Lang doesn't state formally, are recorded in

Lemma V.4.c1. *Let k be a field and f an irreducible polynomial over k. Let α be a root of f in an extension field of k, and let E be a splitting field of f over k. Then the following conditions are equivalent.*

(i) *f is separable.*

(ii) *$k(\alpha)$ is separable over k.*

(iii) *E is separable over k.* □

P.241, last display [=]: An ''*a*'' appears in place of what should be the second ''α''.

P.243, line below first displayed equation [=]: For ''the same field'' read ''the above field''.

P.243, last line [=]: After ''polynomials'' add ''over an algebraic closure of E''.

The idea of this argument is striking: different subextensions of $k(\alpha)$ can be told apart by the different minimal polynomials satisfied by the single element α over each of them.

P.244, end of §V.4 [>]:

Notes on the proof of the Theorem of the Primitive Element.

The preceding proof becomes more understandable if we extract from it the following general result:

Lemma V.4.c2. *A vector space V over an infinite field k cannot be written as the union of finitely many proper subspaces $V_1, \dots, V_n$.*

Proof. Let x be a point of V chosen to *minimize* the number of V_i containing it. We claim that x lies in none of the V_i; this will establish the lemma. Assuming the contrary, let us choose $y \in V$ lying outside of some V_i that contains x, and consider the family of points $x + cy$ $(c \in k)$, which we can picture as a line in V. Any two distinct points of this line clearly span the subspace spanned by x and y; so any subspace of V that does not contain *both* x and y contains *at most* one point of the line. Applying this to each of the finitely many spaces V_i, and noting that because k is infinite, there are infinitely many points on our line, we conclude that some point of our line lies in only those V_i which contain *both* x and y. In particular, it lies in fewer of the V_i than x does, contradicting our choice of x, and completing the proof of the lemma. □

Now in the proof of the Theorem of the Primitive Element, since each subextension of E is in particular a k-subspace, we can replace the second and third paragraphs of the proof by a quick application of the above lemma.

This lemma likewise simplifies the proof of the last statement of the theorem, that every finite separable extension has a primitive element: In the second paragraph of p.244, for each pair of distinct embeddings σ, τ of E in k^{a} let us write $V_{\sigma,\tau}$ for $\{\alpha \in E \mid \sigma\alpha = \tau\alpha\}$. Then by the above lemma, we can find $\alpha \in E$ lying in none of these subspaces. Hence α has degree equal to at least the number of distinct σ, which, since E is separable, is $[E:k]$; so α generates E over k.

Re §V.5. Finite fields.

P.245, long paragraph starting in middle of page [>]: The fact that the roots of $X^{p^n} - X$ form a subring of $\mathbf{F}_p^{\mathrm{a}}$, which Lang proves by detailed calculation, is a consequence of an easily verified and important general fact:

Lemma V.5.c1. *Let $f, g\colon A \to B$ be homomorphisms of rings. Then $\{a \in A \mid f(a) = g(a)\}$ forms a subring $C \subseteq A$.*

The same result is true with ''ring'' replaced by ''A-algebra'', ''A-module'', or ''monoid''.

Moreover, in the ring and monoid cases, if an element a of C has an inverse $a^{-1} \in A$, then $a^{-1} \in C$. Hence in these cases, if A is a field, respectively a group, so is C. □

The subring, subgroup, etc., characterized above is often called the *difference kernel* of f and g, because in the case of modules (or abelian groups), it can be described as the *kernel* of the *difference*-map $f - g$. It is called by categorists the *equalizer* of f and g. If $A = B$, the equalizer of the identity endomorphism and an endomorphism g is called the *fixed subring* (or *fixed subgroup* etc.) of g.

Now the set of roots of $X^{p^n} - X$ in $\mathbf{F}_p^{\mathrm{a}}$ can be described as the fixed subring of the nth power of the Frobenius endomorphism; hence by the above lemma it forms a subfield.

Re §V.6. Inseparable extensions.

P.247, first sentence of §V.6 [>]: Before ''can be omitted'', insert ''except for Proposition 6.1''.

(Also, I disagree with the opinion that this section is more technical, and by implication less interesting than the rest of this chapter.)

P.247, statement of Proposition V.6.1 [=]: Note that ''root of f'' means ''root of f in k^{a}''. This

usage occurs throughout this section; thus, you will find frequent references to the multiplicities of the roots of an irreducible polynomial. Since the roots are in k^{a}, while the polynomials are irreducible over k, there is no contradiction.

Even if one is only interested in separable extensions, this proposition is important, since it shows that every algebraic extension of a field of characteristic 0 – or more generally, every algebraic extension whose degree is not divisible by the characteristic – *is* separable. So Lang's invitation, earlier on this page, to omit §V.6 should have referred only to the material after this proposition.

P.248, end of proof of Proposition V.6.1 [>]: This proof can be made clearer, beginning with the second paragraph on p.248 (''Consider ...''). In the second line, for ''a polynomial'', read ''the polynomial f'''. The result Lang then deduces should in fact be made a lemma: ''*If an irreducible polynomial f is inseparable* (*has multiple roots in its splitting field*), *then the characteristic of the base-field is a prime. An irreducible polynomial f over a field of prime characteristic p is inseparable if and only if it has the form $g(X^p)$.*'' Now consider a general irreducible polynomial f over a field k of characteristic p. Let μ be the largest integer such that we can write $f(X) = h(X^{p^\mu})$. Any factorization of h over our base-field would yield a factorization of f, so h is also irreducible; hence, by the above lemma and our choice of μ, h is separable. Thus, its factorization over k^{a} has the form $h(X) = \prod (X-\beta_i)$, where the β_i are distinct and separable. For each i let us choose $\alpha_i \in k^{\mathrm{a}}$ satisfying ${\alpha_i}^{p^\mu} = \beta_i$. We can now factor f:

$$f(X) \;=\; h(X^{p^\mu}) \;=\; \prod (X^{p^\mu} - \beta_i) \;=\; \prod (X^{p^\mu} - {\alpha_i}^{p^\mu}) \;=\; \prod (X-\alpha_i)^{p^\mu}.$$

This shows that every root has multiplicity p^μ, and has separable p^μth power, completing the proof of the proposition.

We see that, loosely speaking, inseparability is a phenomenon of pth roots in characteristic p. We gave an example of this phenomenon earlier, in our comments on §V.1 (second paragraph of *Hints of what is to come*).

P.250, third line [=]: The argument beginning 8 lines from the bottom of the preceding page (at the word ''Then'') and ending here can be done more simply: by Theorem VI.6.1, every root of the minimal polynomial of α has multiplicity p^μ for some μ; but since the separable degree is 1, there is only one root. So the polynomial has the form $(X-\alpha)^{p^\mu}$. But this equals $X^{p^\mu} - \alpha^{p^\mu}$, hence particular, $\alpha^{p^\mu} \in k$.

P.250, first line of proof of Proposition V.6.5 [=]: ''tower theorem'' should be ''tower property''.

P.250, Corollary V.6.8 and its proof [~]: The same argument shows, more generally, that if K is normal over k, and K_0 is any subextension which is carried into itself by all automorphisms of K, then K_0 is also normal over k.

P.251, line below the diagram [=]: For ''the corollaries of Proposition 6.1'' read ''Theorem 4.1 and Corollary 6.4''.

P.251, first line of Corollary V.6.10 [=]: For ''x^p, $x \in E$'' read ''x^p *where* $x \in E$''.

P.251, first line of Proof of Corollary V.6.10 [=]: Note that from the assumption $E = E^p k$, we get $E = E^p k = (E^p k)^p k = E^{p^2} k^p k = E^{p^2} k$ (the last step because k ''absorbs'' its subfield k^p). Iterating this argument, we get $E = E^{p^m} k$ for all nonnegative m, a fact which will be used two lines later.

P.251, Proposition V.6.11, second line [=]: Rather than turning to the next chapter for the definition of *fixed field*, you can recall Lemma V.5.c1, this Companion, comment to p.245, and note that in that situation if A is a field then C will be a subfield.

P.251, third-from-last line [=]: After ''By definition'' add ''of K^G''.

P.252, first full paragraph [<]: This result can be gotten without reducing to the finite case. Given $\alpha \in K$, let $\alpha_1, \ldots, \alpha_r$ be the distinct roots of the minimal polynomial of α, let $f(X) = \prod (X-\alpha_i)$, and

observe that the coefficients of f will be G-invariant.

P.252, Proof of Corollary V.6.12 [~]: Suppose E is an algebraic extension of k. How do we apply Proposition V.6.11, as Lang asks us to? The point is to first take the normal closure F of E in k^{a}. Then that proposition says that F is a separable extension of a purely inseparable extension; but being perfect, k has no purely inseparable extension, so F is separable, so its subextension E is also separable, as claimed.

Having finished this section, let us note a couple of interesting

Examples involving inseparable extensions.

Both our examples will be based on the following observation:

Let k_0 be a field, p be any prime, and $k_0(u_1, u_2)$ an extension gotten by adjoining independent transcendentals u_1 and u_2 to k_0. Then we see that ${u_1}^p$ and ${u_2}^p$ will also be independent transcendentals over k_0 (since any polynomial relation that they might satisfy could be written as a relation satisfied by u_1 and u_2, and no such relations hold). Moreover, we claim that

$$[k_0(u_1, u_2) : k_0(u_1, {u_2}^p)] \;=\; [k_0(u_1, {u_2}^p) : k_0({u_1}^p, {u_2}^p)] \;=\; p, \qquad \text{whence} \qquad [k_0(u_1, u_2) : k_0({u_1}^p, {u_2}^p)] \;=\; p^2. \tag{c31}$$

Indeed, the degrees in the top line are bounded above by p because each u_i satisfies the polynomial $X^p - {u_i}^p$ over the subfield in question. To also get a lower bound of p, suppose that $1, u_i, \ldots, {u_i}^{p-1}$ were linearly dependent over that field. Taking a linear dependence relation among them, and clearing denominators, we would get an equation $f = 0$ holding in $k_0[u_1, u_2]$. By looking at the residues of the exponents modulo p, we can see that f would be a nonzero polynomial, a contradiction.

The above observations are valid without restriction on the characteristic of k, but in our applications, k_0 will be a field of characteristic p. Our first application will be the construction of

A nonprimitive finite field extension.

Theorem V.4.6 gave a necessary and sufficient condition for a finite extension of a field k to be primitive (i.e., generated over k by a single element), and it was shown that every separable extension satisfied this condition. Moreover, the easiest examples of inseparable extensions, gotten by adjoining to a field a single inseparable element, are ipso facto also primitive.

But suppose we adjoin to a field k of characteristic p pth roots of two elements of k, getting a field $k(a^{1/p}, b^{1/p})$. The Frobenius endomorphism (the pth-power map) carries the elements of k and the two elements $a^{1/p}$ and $b^{1/p}$ into k, hence it carries the whole field they generate into k. This shows that every element of that extension satisfies a polynomial of the form $X^p - c$ ($c \in k$). But $k(a^{1/p}, b^{1/p})$ will generally have degree p^2 over k, so that a primitive element would have to have degree p^2. Hence there is no primitive element.

To complete the example, we need an explicit case of the situation we have just claimed holds ''in general'', i.e., two elements a and b such that $[k(a^{1/p}, b^{1/p}) : k] = p^2$. This is gotten from (c31) by taking any field k_0 of characteristic p, and letting $k = k_0({u_1}^p, {u_2}^p)$, $a = {u_1}^p$, and $b = {u_2}^p$.

According to the Theorem of the Primitive Element, an extension of the sort we have constructed will have infinitely many subextensions. Can we describe an infinite family of such subextensions explicitly? Yes. In general, given an extension $k(a^{1/p}, b^{1/p})$ of degree p^2 in characteristic p, I claim that the subextensions $k(a^{1/p} + c\, b^{1/p})$ as c ranges over the distinct elements of k are all distinct. For each of them has degree $\leq p$ over k, but we see by linear algebra that the compositum of the extensions we get using any two values of c contains both $a^{1/p}$ and $b^{1/p}$, hence is the whole field $k(a^{1/p}, b^{1/p})$. Since this has degree p^2 over k, no two of these extensions of degree p can be equal. (But are we sure that whenever we have such an extension $k(a^{1/p}, b^{1/p})$, there must be infinitely many distinct elements $c \in k$? Yes; because on a finite field k of characteristic p, the pth-power map, being a one-to-one map of a finite set into itself, is a permutation, so in the finite case, there can be no elements a, b which do

not have pth roots in k.)

What I have done above is essentially Lang's Exercise V:L24.

Our second construction will involve

Mixing separable and inseparable extensions.

If E is an algebraic extension of a field k, and F its largest *separable* subextension, then Proposition V.6.6 shows that E is *purely inseparable* over F. Thus, E can be obtained by a separable extension followed by a purely inseparable extension. There is also a largest purely *inseparable* subextension of E, and we might expect that E would be *separable* over this field, giving a decomposition of the reverse sort.

Lang shows in Proposition V.6.11 that this is so if E is normal over k, but warns, in the preceding paragraph, that it is not true in general. Let me show how it can fail. The idea will be to adjoin to k all the roots of a separable polynomial, then adjoin a pth root of just *one* of these. Then the inseparable stage of the extension will have been done in an ''asymmetric'' way, the effect of which *cannot* be obtained by starting with a purely inseparable extension of the base field k.

So let us again look at the situation of (c31) with k_0 a field of characteristic p, and recall that ${u_1}^p$ and ${u_2}^p$ are algebraically independent over k_0. Let us call these two elements t_1 and t_2, let s_1, s_2 denote the *elementary symmetric functions* in this pair of elements, and let us write

$$k \;=\; k_0(s_1, s_2).$$

Thus, k is the field of symmetric rational functions in t_1 and t_2, and $k_0(t_1, t_2)$ is generated over k by the roots t_1, t_2 of the quadratic polynomial $X^2 - s_1 X + s_2$; equivalently, by *one* of the roots, since each can be obtained from the other by subtracting from s_1. We now look at the larger fields gotten by adjoining the pth roots of one or both of t_1 and t_2:

$$E \;=\; k_0(u_1, t_2) \;=\; k(u_1), \qquad L \;=\; k_0(u_1, u_2) \;=\; k(u_1, u_2).$$

Suppose that an element $x \in E$ satisfies $x^p \in k$. From the fact that k consists of elements symmetric in t_1 and t_2, one can see that x will be symmetric in u_1 and u_2. But since x lies in E, u_2 will appear everywhere with exponents divisible by p, hence by symmetry the same will be true of u_1. Thus, x will be a symmetric rational function in t_1 and t_2, hence will lie in k. This means that the largest purely inseparable subextension of E over k is k itself. Since E is not separable over this, we have the desired example.

This construction is delicate. To see this, suppose that, more generally, we adjoin to any field k of characteristic p the roots of an arbitrary quadratic polynomial $X^2 - bX + c$, calling these t_1 and t_2, and then adjoin ${t_1}^{1/p}$, and call the resulting field E. In $k(t_1, t_2)$, we have $t_1 + t_2 = b$, hence in k^{a} we have ${t_1}^{1/p} + {t_2}^{1/p} = b^{1/p}$. (In k^{a} every element has a pth root, and in fields of characteristic p, pth roots are unique when they exist, so the Frobenius endomorphism $x \mapsto x^p$ of k^{a} is bijective, hence is an automorphism. The inverse of this automorphism may be written $x \mapsto x^{1/p}$, and the above equation follows.) Now suppose b and c happened to be chosen so that b is a pth power in k, but c is not; i.e., $b^{1/p} \in k$ but $c^{1/p} \notin k$. When we adjoin ${t_1}^{1/p}$ to $k(t_1, t_2)$, we can get from this the element ${t_2}^{1/p} = b^{1/p} - {t_1}^{1/p}$, and multiplying ${t_1}^{1/p}$ and ${t_2}^{1/p}$ together, we get $c^{1/p} \in E$. Thus, there exists an element $c^{1/p} \in E - k$ purely inseparable over k, which is what the preceding example was aimed at avoiding. In this new example, one finds in fact that E is separable over $k(c^{1/p})$, i.e., this E *is* a separable extension of a purely inseparable extension of k. Essentially the same thing happens if we reverse the hypotheses, and assume $c^{1/p} \in k$ but $b^{1/p} \notin k$.

Chapter VI. Galois Theory.

Re §VI.1. Galois extensions.

P.261, last sentence (continued on next page), ''It is a field because ...'' [=]: This is a special case of Lemma V.5.c1 (this Companion, comment to Lang's p.245).

P.262, line 7 [>]: Lang here mentions several notations for the Galois group of K over k. In the next few pages he will write $G(K/k)$, but later on, he will also use the other notations, rather at random.

P.262, first paragraph of the proof of Theorem VI.1.2 [~]: This repeats the argument used in Proposition V.6.11, and you can easily get the result being proved (that $k = K^G$) from that proposition. Lang doesn't call on the proposition because he made §V.6 optional.

Incidentally, where Lang organizes this proof as a demonstration that the correspondence $H \mapsto K^H$ is one-to-one and onto, I think it is more natural to show that that map, and the map $F \mapsto \mathrm{Gal}(K/F)$ are inverse to one another. Indeed, that is what his proof of one-one-ness and onto-ness reduces to.

P.263, first paragraph [>]: Note that the statement ''H belongs to F'' does *not* mean ''H is the Galois group of F''. Indeed, as an extreme case, the Galois group of our whole extension K is $G(K/F)$, but the subgroup thereof ''belonging to K'' is $\{1\}$.

Incidentally, one also uses the words ''belong to'' in the converse sense, saying ''F is the subfield belonging to the subgroup H'' to mean F is the fixed field of H. (Lang uses it this way at the bottom of p.268.)

P.263, second line of proof of Corollary VI.1.6 [=]: Delete the first occurrence of the word ''finite''.

P.264, first sentence of proof of Theorem VI.1.8 [=]: An easier way to say this would be ''Let $\alpha \in K$ and choose $\sigma_1, \dots, \sigma_k \in G$ so that $\sigma_1\alpha, \dots, \sigma_k\alpha$ are the distinct members of the orbit of α under G''. Since the orbit of an element of a G-set is permuted by every member of G, we can skip the second sentence of the proof and jump directly to the third. (Alternatively, we could drop mention of $\sigma_1, \dots, \sigma_k$ entirely, and begin the proof by writing $\alpha_1, \dots, \alpha_k$ for the orbit of α under G, and using these elements in defining f.)

Four lines after the display in this proof, ''splits in'' should be ''splits into''.

P.264, proof of the Theorem of Artin (VI.1.8) [>]: The first part of this proof again repeats the argument used in Proposition V.6.11.

Note that Theorem VI.1.2 showed that under certain hypotheses, every intermediate field ''belongs to'' some group, while the Theorem of Artin shows that under other hypotheses, every subgroup belongs to some subfield. The limitations of the two results are different: Theorem VI.1.2 requires that we be given a separable normal extension, but needs no assumption of finiteness; the Theorem of Artin requires no separability or normality assumptions, (since it does not start with a field *extension*, but with a single field and a group of automorphisms, and *gives* separability and normality in the conclusion), but finiteness is essential. When we put these results together we get the first assertion of Theorem VI.1.1, a beautiful and powerful bijective correspondence, holding for *finite separable normal* extensions.

P.264, proof of Corollary VI.1.9 [>]: This completes the proof of the first sentence of Theorem VI.1.1, which is what is generally called the Fundamental Theorem of Galois Theory. (The second sentence, to be proved in Theorem VI.1.10, is among the many important auxiliary facts about this correspondence.)

Let us note that although the formulation of that theorem looks about the same from one text to another, there are subtle differences, resulting from the fact that different texts define ''Galois extension'' in different ways. Some authors, like Lang, define this to mean a separable normal algebraic extension; others to mean an algebraic extension E/k such that k is the fixed field of $G(E/k)$. Moreover, one may define ''separable'' either, like Lang, in terms of the number of embeddings in an algebraic closure, or by the condition that the minimal polynomial of every element of the extension have no multiple roots.

Whatever definitions a text gives, it generally supplies the additional results that establish that these definitions are equivalent to the other formulations just mentioned. Thus, Lang's Theorem V.4.3 does this for the concept of separable extension, and Theorem VI.1.8 (Artin's Theorem) for that of normal extension.

So if one takes the conjunction of the Fundamental Theorem of Galois Theory with these supplementary results, this is effectively the same from text to text, even though the statements of the Theorem alone may not be equivalent.

P.264, Remark [>]: I develop the bijective correspondence between subextensions of an *infinite* Galois extension and appropriate subgroups of its Galois group in the comments to Lang's §VI.14 below (though that section is not covered in Math 250A). You can read those comments now (without even looking at that section of Lang), although in the *examples* given, you will have to temporarily take on faith a couple of plausible facts about particular finite extensions, which follow from results in §§VI.3 and VI.8.

Returning to p.264, suppose F is a finite separable extension of a field k, but *not* normal. We know that F can be embedded in a Galois extension K, and, being a field between k and K, it will then ''belong to'' some subgroup $H < G = G(K/k)$. Since the inclusion relations among subfields of K containing k are the reverse of those among subgroups of G, we can say that *subextensions* of F correspond to *subgroups of G that contain H*. Thus, even though the extension field F is not Galois, its subextensions can be described using Galois theory. What about the automorphism group of F over k – can you figure out how to characterize it in terms of G and H? Since it is a finite group of automorphisms of F, the Theorem of Artin says that it is the Galois group of F over a subfield k'. By the Fundamental Theorem of Galois Theory, this k' is the subfield of K belonging to some subgroup of G. Can you characterize that subgroup in terms of H? (If you can't answer these questions now, you might think about them again after finishing this section.)

P.265, slightly below middle, ''There is no avoiding the contravariance ...'' [~]: I am not sure what Lang means by this. I think the point he is making is that if we want to write conjugation by λ in exponential notation, and have it compose correctly for an operation written to the right of its argument, we must define it to take σ to $\lambda^{-1}\sigma\lambda$ rather than $\lambda\sigma\lambda^{-1}$; but then it will take $G(\lambda K/\lambda k)$ to $G(K/k)$, rather than vice versa.

Something group-theorists sometimes do to avoid this sort of problem is to use a left-superscript notation; e.g., writing $\lambda\sigma\lambda^{-1} = {}^{\lambda}\sigma$. (Lang does this on p.69, second and third lines of first Example.)

P.266, statement and proof of Theorem VI.1.12, and diagram on next page [=]: When we adjoin the elements of F to K, not all the automorphisms of K over k may extend to automorphisms of the resulting field over F. Those that do extend form a subgroup, which we know must correspond to some field between k and K. Which field? This theorem says it is $K \cap F$!

If you stop to think about it, you will see that if F has some elements in common with K other than the elements of k, this will indeed prevent automorphisms of K that move those elements from extending to automorphisms of KF over F. The theorem tells us that conversely, F must contain enough of K to lead to whatever failures of automorphisms to extend occur.

The result is sometimes called the *Theorem on Natural Irrationalities*. I think that ''irrationalities'' is an old way of saying ''elements not in the base field'' (generalizing the term from the case where the base field is $\mathbf{Q}$), and the theorem says that F contains those ''irrationalities'' from K that it ''naturally'' ought to.

In the proof, Lang jumps right to proving the last sentence of the theorem. Of the two earlier assertions, the statement that KF is Galois over F holds because normality and separability go over to this extension, while the statement that K is Galois over $K \cap F$ is immediate, since $K \cap F$ is an intermediate field of a Galois extension (Theorem VI.1.2).

P.266, last parenthetical paragraph [~]: Lang has proved the last assertion of the theorem only in the case where K is finite over k (or KF is finite over F), and he now says that topological methods are used in the infinite case. In fact, the infinite case can be deduced from the finite case. Let σ be an automorphism of K over $K \cap F$ which we wish to extend to an automorphism σ_{KF} of KF over F. We note that K is the union of its subextensions K_0 that are finite and normal over k, that each of these subextensions is carried into itself by σ, and that by the finite case, which Lang has proved, the

restriction of σ to each such K_0 extends in a unique way to an automorphism σ_{K_0F} of K_0F over F. Since every element $x \in KF$ may be obtained by field operations from finitely many elements of K and of F, every such x lies in a subextension K_0F, and we may thus define

(c32) $$\sigma_{KF}(x) \;=\; \sigma_{K_0F}(x), \qquad \text{where } x \in K_0F \text{ and } K_0 \text{ is finite over } k.$$

We claim that this definition is independent of our choice of a K_0 for each x, and does indeed give an automorphism. For note that if K_0 and K_1 are two finite normal subextensions of K, then K_0K_1 is also one. The map $\sigma_{K_0K_1F}$ will act on K_0 and K_1 via σ, and will carry K_0F and K_1F into themselves by normality, hence by the uniqueness of σ_{K_0F} and σ_{K_1F}, it will agree with those extensions on their domains. Hence if an element x lies in both K_0F and K_1F, we see that $\sigma_{K_0F}(x) = \sigma_{K_0K_1F}(x) = \sigma_{K_1F}(x)$, so that (c32) gives a well-defined map. That this map is an automorphism can be deduced similarly: Given elements $x \in K_0F$, $y \in K_1F$, these will both lie in K_0K_1F, on which σ_{KF} acts by the automorphism $\sigma_{K_0K_1F}$; thus, we get the required equations $\sigma_{KF}(x+y) = \sigma_{KF}(x) + \sigma_{KF}(y)$, etc..

P.267, diagram, and following line [>]: Lang says, ''It is suggestive to think of opposite sides of a parallelogram being equal''. Note, however, that this only applies under the particular hypotheses of the preceding theorem, which essentially says that *if* one of the lower edges of such a parallelogram is a Galois extension, then the parallel upper edge is also Galois, with the same Galois group. The example marked **Warning** that Lang gives after the next corollary shows that even the equality of degrees between opposite sides can fail if the lower edge is not assumed normal. Moreover, in that example, each of the upper edges of the parallelogram is an extension of degree 2, and every quadratic extension is normal. (Once we adjoin one root ρ of a quadratic polynomial f to a field, the other root is a root of the linear polynomial $f(X)/(X-\rho)$, hence lies in the extension, so the extension is normal.) So in this diagram, the *top* two edges are normal, yet opposite edges are still not in any sense ''equal''. Incidentally, in that ''Warning'', all occurrences of the symbol ''E'' should be changed to ''K'' (six occurrences in the last six lines) for conformity with the preceding material.

P.269, last phrase of §VI.1 [=]: ''Notation should be functorial with respect to the ideas!'' was a slogan of Lang's. It is not clear to me exactly how he was using the word ''functorial''. One way of looking at it is that when one has a functor between categories, not only does each object of the domain category yield an object of the codomain category, but morphisms yield morphisms, in a way that elegantly matches the structure of the domain category, and that Lang was implying that notation should similarly elegantly match ideas.

Note, also, that in pre-category-theoretic usage, when one had a construction of a new object $F(X)$ from an old object X, and one (eventually) recognized that every map $a\colon X \to Y$ induced a map $F(X) \to F(Y)$, one generally denoted that map by an ad hoc symbol such as a^*; but with the realization that the constructions on objects and on morphisms form one entity, a ''functor'', we now write $F(a)\colon F(X) \to F(Y)$. Lang may have been saying that notation should reflect ideas in something like the way that functorial notation does. His use of k^{a}, k^{s} and k^{ab} rather than arbitrary bars and tildas does this.

P.269, end of §VI.1 [>]:

Galois connections.

The Fundamental Theorem of Galois Theory is an example of a very general type of mathematical situation: One has two sets S and T, and a certain binary *relation* on them, i.e., a subset $R \subseteq S \times T$. (In our case, S is an extension field K of a field k, T is the set $G(K/k)$, and R is the ''fixed-element'' relation, $\{(x, \sigma) \in K \times G(K/k) \mid \sigma(x) = x\}$.) For every subset $X \subseteq S$ one defines $X^* \subseteq T$ to be $\{t \in T \mid (\forall\, s \in X)\ (s,t) \in R\}$, and likewise for every subset $Y \subseteq T$ one defines $Y^* \subseteq S$ to be $\{s \in S \mid$

$(\forall\, t \in Y)\ (s,t) \in R\}$. These operators constitute what is called a *Galois connection* between S and T, the general concept being named after the important special case we have been studying.

There are a number of basic results about Galois connections that are trivial to prove, but worth knowing, to keep from having to rediscover them in case after case! In particular, (i) $X \subseteq X' \Rightarrow X* \supseteq X'*$; (ii) the operators $**$ from subsets of S to subsets of S, and from subsets of T to subsets of T, are *closure operators* on S and T respectively (as defined in this Companion, comments on §III.1); and (iii) the operators $*$ give a bijective inclusion-reversing correspondence between closed subsets of S and closed subsets of T with respect to these operators. (These and other results are proved, and numerous examples discussed, in §5.5 of the 245 notes.)

What *cannot* be gotten from the general theory of Galois connections are *characterizations* of the closed sets in a given case – for example, the result that when K is a finite separable normal extension of k and R is the ''fixed element'' relation between elements of K and elements of $G(K/k)$, then the closed subsets of K are precisely all intermediate fields, and the closed subsets of $G(K/k)$ are all its subgroups. For each Galois connection, this problem must be investigated using the methods of the area of mathematics in which it arises.

Re §VI.2. Examples and applications.

P.269, beginning of §VI.2 [<]: There is a fundamental class of examples which Lang has given us the background for, but which he fails to note here! Since it is of a slightly different sort from the examples he gives, let us note it before you start this section.

We have seen that if K is a field of characteristic p, then the pth power operation is an endomorphism of K, called the Frobenius endomorphism, whose fixed field is precisely the subfield of p elements (since $X^p - X$ can have at most p roots). If K is a finite field, say $\mathbf{F}_{p^n}$ (in the notation of §V.5), then this endomorphism will necessarily be an automorphism, and be of finite order; hence by the Theorem of Artin, the group it generates will be the full automorphism group of $\mathbf{F}_{p^n}$ over its fixed field $\mathbf{F}_p$. Thus, each field $\mathbf{F}_{p^n}$ is a Galois extension of $\mathbf{F}_p$, with cyclic Galois group, which must have order $[\mathbf{F}_{p^n} : \mathbf{F}_p] = n$.

One can also see (in more than one way) that for every m dividing n, the subgroup generated by the mth power of the Frobenius map corresponds to the subfield $\mathbf{F}_{p^m} \subseteq \mathbf{F}_{p^n}$.

P.269, first paragraph of §VI.2 [>]: There are some general observations about this situation that we should make here for use in the examples to follow. First, f will be irreducible if and only if its Galois group G is *transitive* on the set of roots. Indeed, if we have a factor g of f in $k[X]$, then g is fixed under G, hence the set of its roots must be carried into itself by this group; and, conversely, given a subset A of the set of roots of f that is carried into itself by G, we see that $\prod_{\alpha \in A} (X - \alpha)$ is a factor of f in $K[X]$ which is invariant under G, hence lies in $k[X]$, i.e., a factor of f in the latter ring. Thus, f has no proper factor if and only if G acts transitively on the set of its roots. Finally, note that if we adjoin to k one of the roots of f, say α_1, then the subgroup of G belonging to the resulting subextension $k(\alpha_1) \subseteq K$ will be the group of elements of G fixing α_1. (Can you state a precise proof of this?) If, with Lang, we regard G as a subgroup of S_n, i.e., as acting formally on $\{1, \dots, n\}$ rather than on $\{\alpha_1, \dots, \alpha_n\}$, then this becomes the isotropy subgroup of the element 1.

P.269, third from last line [=]: For ''This comes from completing the square and'' read ''This may be seen by taking the irreducible polynomial of any generator of K over k, and completing the square, or equivalently,''.

We remark that in the preceding paragraph, Lang introduced the assumption $\mathrm{char}\ k \neq 2$ because for the characteristic 2 case, a different method is needed, not taught in ''elementary school'', which we will see at the end of §VI.7. For similar reasons he will assume $\mathrm{char}\ k \neq 2, 3$ in Example 2.

P.270, line 10 [=]: ''In the second case $k(\alpha)$ is not normal over k'': This is because the corresponding group is the stabilizer of $1 \in \{1, 2, 3\}$ in S_3, which is not a normal subgroup. In contrast, the stabilizer of 1 in A_3 is normal therein; equivalently, it fixes all members of the orbit of 1 under

that group.

P.270, line below display ''Δ = ...'' [>]: Lang writes ''The set of σ in G which leave δ fixed is precisely the set of even permutations''. You can probably convince yourself that this is true, but let us develop formally the corresponding result for equations of arbitrary degree. Note that D and δ are the polynomials that Lang introduced on pp.192-193 (though he is writing D here for what he called D_f there, and Δ here for what he called D there, i.e., the result of substituting the coefficients of a particular polynomial for the arguments of D. I will instead distinguish these explicitly as $D(s_1, \dots, s_n)$ and $D(a_1, \dots, a_n)$.)

Lemma VI.2.c1. *Let n be a positive integer, and let $D \in \mathbf{Z}[X_1, \dots, X_n]$ be the* discriminant polynomial, *that is, the unique polynomial such that in the polynomial ring $\mathbf{Z}[t_1, \dots, t_n]$, if one writes s_i for the ith elementary symmetric polynomial in the t_i, one has $\prod_{i<j} (t_i - t_j)^2 = D(s_1, \dots, s_n)$.*

Then for any field k not of characteristic 2 and any separable monic polynomial $f = X^n + a_{n-1}X^{n-1} + \dots + a_0$ of degree n over k, if we denote by E the splitting field of f, and by $\alpha_1, \dots, \alpha_n$ the roots of f in E, then the following three conditions are equivalent:

(a) *$G(E/k)$, regarded as a group of permutations of $\alpha_1, \dots, \alpha_n$, lies in the alternating group A_n (i.e., acts by even permutations on these roots).*

(b) *The element $\delta(\alpha) = \prod_{i<j} (\alpha_i - \alpha_j)$ of E lies in k.*

(c) *The element $D(a_{n-1}, \dots, a_0) \in k$ is a square in k.*

Hence, the fixed field of the group of elements of $G(E/k)$ which act by even permutations on the roots of f is $k(\delta(\alpha)) = k(D(a_{n-1}, \dots, a_0)^{1/2})$.

Proof. It is easy to see that the automorphism of $\mathbf{Z}[t_1, \dots, t_n]$ that acts on the subscripts of the t's by a permutation of the form $(i, i+1)$ takes $\delta(t) = \prod_{i<j} (t_i - t_j)$ to $-\delta(t)$. Since an *odd* (respectively *even*) permutation can be characterized as a permutation which can be written as the product of an odd (respectively even) number of permutations of the form $(i, i+1)$, we see that the action of any odd permutation on $\mathbf{Z}[t_1, \dots, t_n]$ will send $\delta(t)$ to its negative, and any even permutation will send it to itself.

Now if an automorphism θ of E acts by a permutation π_θ on the roots $\alpha_1, \dots, \alpha_n$ of f, then it is easy to see that the map $\mathbf{Z}[t_1, \dots, t_n] \to E$ carrying t_i to α_i makes a commuting square with the automorphism θ of the field, and the automorphism of that polynomial ring acting by π_θ on the subscripts of the indeterminates. It follows that θ will send $\delta(\alpha)$ to itself if π_θ is even, and to its negative if π_θ is odd. Hence, if all members of $G(E/k)$ act by even permutations on $\alpha_1, \dots, \alpha_n$, then $\delta(\alpha)$ is fixed under that group, hence belongs to k, proving (a)$\Rightarrow$(b); and conversely, if this element belongs to k, then all members of the Galois group fix it, hence act by even permutations; so (b)$\Rightarrow$(a).

It is also clear that $D(a_{n-1}, \dots, a_0) = \delta(\alpha)^2$; hence (b)$\Rightarrow$(c). Since the equation $X^2 - D(a_{n-1}, \dots, a_0) = 0$ can have at most two roots in E, its only roots are $\pm\delta(\alpha)$, so (c)$\Rightarrow$(b).

The last assertion is easily deduced from the above results. □

Thus Lang is essentially applying the above lemma with $n = 3$, using the formula for $\Delta = D(0, a, b)$ obtained in Chapter IV (p.193). Another important application of the above lemma will be made in Exercise VI.3:1.

Let us record one other important fact about the discriminant. You should find it straightforward to prove, by the same approach as above,

Lemma VI.2.c2. *Let n and D be as in the preceding lemma, let k be* any *field, and let $f = X^n + a_{n-1}X^{n-1} + \dots + a_0$ be any monic polynomial of degree n over k. Then f is inseparable* (*has at least one multiple root in k^a*) *if and only if $D(a_{n-1}, \dots, a_0) = 0$.* □

P.271, the two lines after second display (''The field ...'') [=]: Replace these two lines by, ''Since α is real, $i \notin \mathbf{Q}(\alpha)$. But i satisfies the polynomial $X^2 + 1$, hence it has''. And the next sentence should begin with ''So''.

P.271, next-to-last line of text, ''This gives us the structure of G'' [>]: It is easy to see that the four roots of $X^4 - 2$ in the complex plane form the vertices of a square, and that the group Lang shows to be the Galois group of this equation acts on these roots as the full symmetry group of that square! You should be able to determine the field extension corresponding to each subgroup in the diagram at the bottom of that page. In that diagram, the symbols $\langle \ldots \rangle$ should more logically be $\{ \ldots \}$, since Lang is listing all the elements of each subgroup, rather than just a generating set. (A wide class of extensions having this same Galois group, or subgroups thereof, is examined in Exercises VI.2:1 and VI.2:2.)

P.272, the two lines after second display [>]: Since F is the fixed field of a group of order $n!$, K has degree $n!$ over it, hence it has degree at least $n!$ over $k(s_1, \ldots, s_n)$, with strict inequality if this field is strictly smaller that F. But since K is generated over that subfield by the roots of the polynomial f which has coefficients in that subfield, the degree of K over that subfield is also at most $n!$. Hence that subfield is exactly F.

Here, as in the proof of Artin's Theorem (p.264), and of the Primitive Element Theorem (p.243, argument at the bottom of the page), the fact that a subfield contains the coefficients of one or more irreducible polynomials helps us ''pin down'' which subfield it is.

P.272, last paragraph of Example 4 [=]: The theorem referred to in the first sentence of the paragraph is one we have not had, but the consequence that there exists a Galois extension E of $\mathbf{Q}$ such that $G(E/\mathbf{Q}) \cong S_n$ is proved for the special case of n prime in Example 6 on the next page, and for the general case in Exercise VI:L14, p.322. Concerning the open question Lang then discusses, note that if G is a finite group, we can embed it in a symmetric group S_n. If we let E be an extension of $\mathbf{Q}$ having S_n as Galois group, as above, and F the intermediate field corresponding to the given group $G \subseteq S_n$, then $G(E/F) \cong G$. Thus, every finite group is the Galois group of a Galois extension E/F such that E and F are both finite extensions of $\mathbf{Q}$. The open question is whether one can always do this with $F = \mathbf{Q}$. The final two sentences, about ''specializing parameters'' is a technical point which you may ignore.

P.272, beginning of Example 5 [~]: Lang mentions that $\mathbf{R}$ is ''an ordered field''. This means a field with a total ordering that satisfies certain of the properties of the ordering of the real numbers. Here all you need to know is that you will be using in the proof some of the familiar properties of the order relation on the reals. The general definition of an ordered field may be found in §XI.1.

Toward the bottom of the page, Lang shows that an arbitrary complex number $a + bi$ has a square root. To see how one could discover the formula he gives, consider the problem of solving for c and d the equation $a + bi = (c + di)^2$. Expanding the right-hand side and equating real and imaginary parts, one gets a value for $c^2 - d^2$ and a value for cd. Squaring the latter and multiplying by -1, we see that we have values for the sum and product of c^2 and $-d^2$, i.e., the elementary symmetric functions in these two real numbers. Hence we can find c^2 and $-d^2$ by the quadratic formula. (Obviously, c^2 will be the larger root, and $-d^2$ the smaller.) The fact that the reals are an ordered field is used to show that the number whose square root one has to take in using this formula is nonnegative, and that the resulting values for c^2 and $-d^2$ are nonnegative and nonpositive respectively, hence that the values for c^2 and d^2 are both nonnegative. If one arbitrarily chooses a square root of the value obtained for c^2, and calls this c, then, depending on one's choice of a square root d of d^2, one finds that $(c + di)^2$ comes out as $a \pm bi$. In particular, one choice gives the desired value, $a + bi$.

P.273, beginning of second paragraph [=]: Change ''We now see'' to ''Recall''.

P.273, end of Example 5 [>]: The argument Lang has given may be summarized as follows. The fact that every finite group has a p-Sylow subgroup tells us that a group whose order is not a power of p has a subgroup whose index is not divisible by p; while we also know that every nontrivial group whose order *is* a power of p has a subgroup of index exactly p. Using the Fundamental Theorem of Galois

Theory, these statements translate to say that if a field has an extension of degree not a power of p, it has a nontrivial extension of degree relatively prime to p, while if it has a nontrivial extension of degree a power of p, it has an extension of degree exactly p. Now taking $p = 2$, we observe that if the complex numbers had an extension of degree not a power of 2, this would also be an extension of $\mathbf{R}$ of degree not a power of 2, hence by the above observations, $\mathbf{R}$ would have a nontrivial extension of degree not divisible by 2, and the minimal polynomial of some element thereof would be a real polynomial of odd degree having no real root, a contradiction. On the other hand, if $\mathbf{C}$ had a proper extension of degree a power of 2, then by the above observations, it would have an extension of degree exactly 2. But using the quadratic formula, and the existence of square roots of all complex numbers, we know that every quadratic polynomial over $\mathbf{C}$ splits, so this is also impossible. Thus $\mathbf{C}$ has *no* nontrivial finite extensions.

P.273, Example 6, sentence beginning ''When-'' [=]: For ''this group'' read ''a subgroup of S_n'', and for ''this representation of G'' read ''the Galois group of f, regarded as a group of permutations of the roots of f''.

P.274, through end of §VI.2 [<]: You can skip these last three examples, which assume material in later chapters, or not in the book.

If you do read Example 7, trusting the truth of the result Lang quotes in italics, you might see Exercise VI.2:4 for a clarification of an argument Lang gives later in that example regarding the Galois group of $X^2 - X - 1 \pmod 2$.

If you have seen the topological concept of covering space, and had an undergraduate course in complex analysis, you can read in place of (or before) Example 9 the following more elementary notes on the relation between

Galois theory and Riemann surfaces (sketch, optional).

The Riemann sphere $\mathbf{C} \cup \{\infty\}$ (which Lang calls the Gauss sphere) is defined to consist of the complex plane with one ''point at infinity'' adjoined. (A picture can be found in most complex analysis texts.) Every member of $\mathbf{C}(t)$ can be regarded as a meromorphic function on this object; e.g., t is a function with a ''pole'' at infinity, but holomorphic everywhere else; $1/(t-1)$ has a pole at 1, but is holomorphic everywhere else (including ∞).

Now consider the extension field $\mathbf{C}(t^{1/2})$. At each point of the Riemann sphere other than 0 and ∞, there are two values that can be given to $t^{1/2}$, hence a general rational function of this element will be double-valued. To get around this problem, one regards elements of this field as meromorphic functions, not on the Riemann sphere, but on a ''two-sheeted branched covering of that sphere, with branch-points at 0 and ∞''. This means, essentially, a space X with a map into the Riemann sphere, such that if we remove those two points from the sphere, and their inverse images from X, what we are left with is a two-to-one covering map. (Certain conditions are also assumed on how the map behaves at the branch points, which I won't go into.)

This particular branched covering of the Riemann sphere can itself be identified with the Riemann sphere itself, by using the parameter $t^{1/2}$ in place of t; i.e., it is the Riemann sphere ''wrapped twice around itself''. But if we take an algebraic extension of $\mathbf{C}(t)$ that cannot be generated over $\mathbf{C}$ by a single element, for example, $\mathbf{C}(t, (t^3 - t)^{1/2})$, then the corresponding branched covering, when ''unwrapped'', is not a sphere, but a torus, or a more complicated surface. The objects one gets in this way are called ''compact Riemann surfaces''. Given a branched covering of the Riemann sphere by such a surface (or of one such surface by another), it can be shown that the Galois group of the corresponding field extension acts by self-homeomorphisms of the covering space, that respect the covering map, and hence permute the ''sheets''. In other words, the Galois group becomes the covering group of our topological covering!

(The Berkeley courses that consider Riemann surfaces are Math 205, 241 and 255.)

(Lang's Example 9 skips past the finite extensions of $\mathbf{C}(t)$ discussed above, and looks at what is, in

effect, the *compositum* (union) of all such finite extensions whose branch points lie in a given finite set $\{P_1, \dots, P_{n+1}\}$. (On line 5 he has P_{n-1} instead of P_{n+1} by mistake.) As discussed in the note to §VI.14 below, the Galois group of such an infinite extension can be described in terms the Galois groups of the finite extensions of which it is a union.)

Re §VI.3. Roots of unity.

P.277, first paragraph [>]: In any field, a *primitive nth root of unity* means an element of multiplicative order n. A field may or may not have such elements; for example, $\mathbf{Q}$ has a primitive square root of unity but no primitive nth root of unity for higher n, though its extension $\mathbf{C}$ has primitive nth roots of unity for all n. $\mathbf{F}_7$ has primitive square, cube and sixth roots of unity (hence so does any field of characteristic 7); it has no primitive twelfth root of unity, but $\mathbf{F}_{7^2}$ does. (Do you see why?) No field of characteristic p has a primitive pth root of unity (do you see why not?), hence no such field has a primitive nth root of unity for any n divisible by p.

In this paragraph, Lang assumes n not divisible by the characteristic of k, and works in k^{a} (so that what he is looking for could be described as nth roots of unity *over* k, rather than *in* k). In this context, he shows that the group $\boldsymbol{\mu}_n$ is cyclic of order n (do you see why its order will be exactly n?), hence, a generator of this group will indeed be a primitive nth root of unity.

P.277, line after third display, ''Hence $\sigma\zeta = \zeta^i$,'' [>]: Note that since every nth root of unity has the form ζ^r for some r, σ will act on these roots by $\sigma(\zeta^r) = (\sigma\zeta)^r = \zeta^{ir} = (\zeta^r)^i$. In particular, the same integer i for which $\sigma\zeta = \zeta^i$ also satisfies $\sigma\zeta' = \zeta'^i$ for all primitive nth roots of unity ζ'. This is why Lang can write $i = i(\sigma)$, rather than $i = i(\sigma, \zeta)$.

P.277, fifth line from bottom [=]: The function $\varphi(n)$, called the *Euler phi function*, was defined on p.94.

P.277, third line from bottom [=]: Here is where Lang starts using variant notations for the Galois group (mentioned on p.262, end of second paragraph). Also, on this and the next line there are three occurrences of κ that should be k as in the beginning of this line.

P.278, statement of Theorem VI.3.1 [>]: To see what is needed to prove this result, consider any field k, and any positive integer n not divisible by the characteristic of k. Suppose we write $X^n - 1$ as a product of irreducibles $g_1 \dots g_r$ in $k[X]$, and look at the roots of each factor in a splitting field of $X^n - 1$. Since all the roots of an irreducible polynomial over k satisfy precisely the same equations over k, we can say that for each i, all roots of g_i have the same multiplicative order. In particular, the $\varphi(n)$ primitive nth roots of unity, which are the roots of order n, will constitute the set of all roots of some subfamily of these factors. From these observations, we can see that Theorem VI.3.1 is equivalent to the statement that when $k = \mathbf{Q}$, all the primitive nth roots of unity are roots of a *single* irreducible factor of $X^n - 1$. (Lang will make this translation at the bottom of p.279; but in fact, it is implicit in his proof.)

P.278, proof of Theorem VI.3.1 [~]: The idea of this proof can be best understood if we ask ourselves how much of it works if we replace $\mathbf{Q}$ by a more general field k. So let k be any field of characteristic not dividing n, and let us look at the primitive nth roots of unity in a splitting field of $X^n - 1$. These roots are permuted transitively by the operations of raising to various positive integer powers m relatively prime to n. Let us ask which of these values of m have the property that exponentiation by m preserves (permutes) the set of roots of every irreducible factor of $X^n - 1$. If you examine the method of proof of the theorem, you will see that it shows that if k is the field of fractions of some UFD A which can be mapped homomorphically to a field $\mathbf{Z}/p\mathbf{Z}$ (p relatively prime to n), then exponentiation by p permutes the roots of each irreducible factor of $X^n - 1$. The key facts from which this follows are: (i) exponentiation by p is an automorphism of any algebraic extension of $\mathbf{Z}/p\mathbf{Z}$, hence permutes the roots of any polynomial over $\mathbf{Z}/p\mathbf{Z}$, (ii) the property that raising to the pth power sends roots of an irreducible polynomial g_i to roots of a polynomial g_j can be written as a divisibility relation, $g_i(X) \mid g_j(X^p)$, and so is preserved on going to homomorphic images (even if the irreducibility of g_i is lost), and (iii) $X^n - 1$ is separable over $\mathbf{Z}/p\mathbf{Z}$, so the distinct factors g_j of $X^n - 1$ over $k[X]$, when

taken in $A[X]$ and then mapped into $\mathbf{Z}/p\mathbf{Z}[X]$, remain relatively prime, hence the various $g_j(X^p)$ are also relatively prime, hence $g_i(X)$ cannot divide more than one $g_j(X^p)$.

Returning to the special case $k = \mathbf{Q}$, note that the UFD $\mathbf{Z}$ has homomorphisms into all fields $\mathbf{Z}/p\mathbf{Z}$. Hence exponentiation by every positive integer prime to n permutes the roots of each factor of $X^n - 1$ over $\mathbf{Q}$. But every primitive nth root of unity can be obtained from any other by raising it to a positive integer power; thus, all the primitive nth roots of unity are roots of a single factor of $X^n - 1$.

P.278, proof of Corollary VI.3.2 [<]: In preparation for this proof, let us make some easy observations on the function φ. If m and n are relatively prime, we know that $Z_{mn} \cong Z_m \times Z_n$ (Proposition I.4.3(v)) and that the cyclic generators of this group correspond to the pairs (x, y) where x and y are generators of the respective factors. It follows that

For m and n relatively prime, $\varphi(mn) = \varphi(m)\varphi(n)$.

A number-theoretic function with this property is called *multiplicative*. (Note that the word does *not* mean, as one might imagine, that the above equation holds for *all* m and n!) To evaluate such a function on all positive integers, it suffices to know its values on powers of primes. As noted on p.96, $\varphi(p^r) = (p-1)p^{r-1}$.

P.278, line following next-to-last display (''Our assertion ...'') [=]: Add at the end of the line, ''and Corollary 1.13''.

P.279, line following second display [=]: Note that the periods (i.e., orders) of nth roots of unity will be the divisors of n, and that if d is a divisor of n, then the dth roots of unity are a subset of the nth roots of unity. Hence the product defining $\Phi_d(X)$ is indeed a function of d alone, and not of n, even though we are looking at it as the product of a subset of the factors of $X^n - 1$.

P.279, 5th line from bottom [=]: ''Cyclotomic'' means ''circle-cutting'', the idea being that the nth roots of unity divide the unit circle of the complex plane into n equal arcs.

P.280, assertions **1-5** [<]: You should be able to get assertions 3 and 4 as immediate consequences of the definition of Φ_n. Assertion 2 and the second part of assertion 1 are applications of assertion 4, while the first part of assertion 1 has already been proved.

Assertion 5, as well as some subsequent discussion on this page, involve *Möbius inversion*. You will not be responsible for this (or any of the material from assertion 5 through the first paragraph on p.281), but the method is a general and important one, so I will sketch the idea.

Suppose one has a function α (not assumed a homomorphism) from the positive integers to an abelian group A (which we will write *multiplicatively*), and another such function β related to α by the formula $\beta(n) = \prod_{d|n} \alpha(d)$, the product being taken over all positive integer divisors of n, including 1 and n. It is not hard to see that the values of α can be recovered from those of β. Indeed, assuming inductively that one has found $\alpha(m)$ for all proper divisors m of n, one can get $\alpha(n)$ by dividing $\beta(n)$ by the product of those values. So in fact, $\alpha(n)$ is expressible, using the group operation of A, in terms of the values of $\beta(d)$ as d ranges over the divisors of n. The exact formula can be shown to be $\alpha(n) = \prod_{de=n} \beta(d)^{\mu(e)}$, where μ, the *Möbius function*, is defined by the formula Lang gives on this page. He applies this result with $A = Q(X)^*$, $\alpha(n) = \Phi_n(X)$ and $\beta(n) = X^n - 1$. (Lang refers to this technique, in the last paragraph of the page, as ''inverting by the general formalism of convolutions''. This is the same ''convolution'' as in the definition of the multiplication operation of a monoid algebra, but to explain the connection would require a lengthy digression, in which we would turn each of the functions α and β above into a formal power series in infinitely many indeterminates; so I will not go into it here. Some of the ideas are developed in Exercises II:L12, p.116, and V:L21-22, p.254.)

P.281, 2nd display [=]: To make this calculation clearer, I would move the fraction to between the n-term sum and the final 0. In any case, the argument is that the sum equals that fraction by the general formula for $(X^n - 1)/(X - 1)$, and the fraction is zero because its numerator is zero. If ζ is a *primitive* nth root of unity, another way to see this same result is by noting that the sum is the sum of the roots of $X^n - 1$, hence equals the coefficient of X^{n-1} in that polynomial, which is zero. In the field of complex

numbers, we can also see the result geometrically: The powers of ζ are the vertices of a regular n-gon, which has center of gravity at the origin, hence 0 is the average of the vertices as complex numbers, hence they sum to 0. Finally, if ζ is a not necessarily primitive nth root of unity, it will still be a primitive dth root for some $d|n$, where $d \neq 1$ by assumption. We see that the sum in question will be the sum of the distinct powers of ζ each repeated n/d times. Hence by the ''primitive root'' case, this sum is zero.

P.281, statement of Theorem VI.3.3 [=]: Here, as in the preceding discussion, p is an *odd* prime (a prime $\neq 2$). Note that from the way S is defined, if we apply an automorphism that sends ζ to ζ^μ, then if μ is a square mod p, S will be unchanged, while if μ is a non-square, S will change to $-S$. Thus S^2 is invariant under the whole Galois group, and taking our base-field to be $\mathbf{Q}$, we conclude that S^2 must be a rational number, whose square root belongs to the indicated cyclotomic extension. The point of the theorem is to establish which rational number it is. The answer, we are told, is $\pm p$, with sign depending on whether -1 is a square mod p.

Let me sketch how one might come up with this answer and its proof. Squaring the formula defining S, we get the expression in the first display on the next page. What would be an enlightening way of grouping the terms? We might try to group them by the sum $\mu+\nu$. This looks nice in terms of the exponent of ζ; but it is not clear how the quadratic symbol will behave as μ and ν vary with constant sum. If we tried to group them according to μ or ν alone, we would just be undoing our squaring operation. Suppose, then, that we group them by the *ratio* of μ to ν, calling that ratio λ, so that ν becomes $\lambda\mu$. Then the ''numerator'' in the quadratic symbol becomes $\lambda\mu^2$, and since the quadratic symbol is unchanged under multiplying by a square, it will be the same in each group of terms. This gives us the right-hand side of the top line of Lang's second display; except that he has recycled the symbol ν with a new meaning, where I have introduced a new symbol λ; and I would write double summation, the outer Σ being over λ, which ranges over all nonzero residues modulo p. The quadratic symbol would be written occurring right after that Σ, and the inner summation would be $\Sigma_\mu\ \zeta^{\mu(\lambda+1)}$, where μ also ranges over nonzero residues modulo p.

Now evaluation the inner sums is complicated slightly by the exclusion of the residue $\mu = 0$. Note that the sum of the excluded terms is just the sum of all the quadratic symbols, multiplied by $\zeta^0 = 1$. Since the quadratic symbol takes on the values $+1$ and -1 equally often, the sum of those excluded terms is 0, and we can put them in. We then see that the summation $\Sigma_\mu\ \zeta^{\mu(\lambda+1)}$ is *almost* always just the sum of all powers of ζ, which is 0 (since it is unchanged under multiplying by ζ). The one exception is when $\lambda = 1$, where we are summing $\zeta^0 = 1$ p times, and the sum is p. Bringing in the coefficient $\left(\frac{\lambda}{p}\right)$, that term comes to $\left(\frac{-1}{p}\right) p$, as claimed.

(Incidentally, note that ''μ'' in this proof is just an index representing an integer modulo p; not to be confused with the group ''$\boldsymbol{\mu}_n$'' of p.277 nor the Möbius function $\mu(n)$ of p.280.)

Re §VI.4. Linear independence of characters.

P.283, top paragraph [>]: Given any field K and any set S, we know that the set K^S of K-valued functions on S becomes a K-vector-space if addition and scalar multiplication are defined componentwise. Lang is here regarding characters $G \to K$ as forming a *subset* of this vector space, and considering whether this subset is linearly dependent.

In the simplest case, where $G = \mathbf{N}$, the monoid of nonnegative integers, characters $G \to K$ are sequences of powers $(1, \alpha, \alpha^2, \ldots)$. You may have been aware that distinct such sequences are always linearly independent. (In fact, the formula for the Vandermonde determinant shows that given n such sequences, the length-n vectors formed from their first n terms are linearly independent. Though this is not essential to the present reading, it is a fact of general importance in mathematics, and I will review it below.) The result of this section shows that the same is true for homomorphisms $G \to K$ for an arbitrary monoid G.

The Vandermonde determinant (sketch, optional).

The Vandermonde determinant is defined as the polynomial in n indeterminates $X_1, \dots, X_n$ given by the determinant of the $n \times n$ matrix whose jth column is composed of the $j-1$st powers of the indeterminates:

(c33) $$\det((X_i^{\,j-1}))_{i,j=1,\dots,n},$$

(or more generally, the ring element resulting from the substitution of some n values x_i for the indeterminates X_i in this expression). I claim that (c33) equals

(c34) $$\prod_{i>j} (X_i - X_j).$$

The first step is to show that (c33) is divisible by all the factors shown in (c34). Intuitively, it is divisible by each term $X_i - X_j$ because if one sets X_i and X_j equal, then (c33) becomes the determinant of a matrix whose ith and jth rows are equal, and thus goes to zero. To formalize the concept of setting X_i and X_j equal, let us map the polynomial ring in $X_1, \dots, X_n$ to its factor-ring by the ideal generated by any difference $X_i - X_j$. Then the image of (c33) in that factor-ring is the determinant of a matrix over that ring with two equal rows, and is thus zero; hence (c33) must lie in the ideal we have factored out. Since this is true for all $i < j$, and since the distinct terms $X_i - X_j$ $(i > j)$ are relatively prime, (c33) is divisible by (c34). Moreover, (c33) and (c34) are of the same degree, $0+1+\dots+(n-1)$, hence they can only differ by a scalar factor. To determine this scalar, we look at the coefficient of the monomial

$$X_2\, X_3^2 \dots X_{n-1}^{n-2}\, X_n^{n-1}$$

in (c33) and in (c34). This monomial arises exactly once in the expansion of each (as the diagonal product in (c33), and as the product of the first term of each factor in (c34)), hence it has coefficient 1 in each, hence the scalar in question is 1.

Now note that whenever we map our polynomial ring into a field so as to take $X_1, \dots, X_n$ to *distinct* values, $x_1, \dots, x_n$, (c34) goes to a nonzero field element; hence the matrix whose determinant is taken in (c33) must have linearly independent rows. This shows, as we claimed earlier, that any n distinct sequences of the form $(1, \alpha, \alpha^2, \dots, \alpha^{n-1})$ in a field are linearly independent.

P.284, before the line preceding Corollary VI.4.2 [<]: At this point read Corollary VI.5.4, p.286, which for some reason was put in the wrong section. (That corollary is a slight generalization of the observation you have just read on p.284 of Lang)

P.284, Corollary VI.4.2 [~]: This is the result we obtained in a more precise form using the Vandermonde determinant.

P.284, end of §VI.4 [>]. Here are two tangential notes related to the material of this section.

Further observations on linear independence of characters (optional).

The theorem on Linear Independence of Characters was proved in preparation for an application that will be made in the next section; but let us note that it can also be used (and is, in some textbooks) in place of one of the steps in Lang's development of the Fundamental Theorem of Galois Theory; namely, to show that the order of the automorphism group of a finite extension is at most the degree of that extension, or more generally, that the number of embeddings over k of an extension field E in an algebraic closure k^{a} is at most $[E:k]$ (p.240, top line). This was done in Lang by counting roots of polynomials; the alternative argument is as follows: An embedding of E in k^{a} is, in particular, a k^{a}-valued character on the multiplicative monoid of E; so the set of such embeddings is linearly independent. In addition to being a character, such an embedding is a k-linear map $E \to k^{\mathrm{a}}$, and such a map is determined by what it does on elements of a k-basis of E. Thus, the k^{a}-dimension of the space of all k-linear maps $E \to k^{\mathrm{a}}$ is $[E:k]$. Combining these two observations, we see that this number is an upper bound on the number of

embeddings of E in k^{a} over k.

Turning to the way linear independence of characters is proved, it is interesting that essentially the same the argument is used in getting several important mathematical results: one assumes the existence of a linear relation among certain elements, takes such a relation that involves the fewest terms; then transforms it in some way, and shows that by taking a linear combination of the original relation and the transformed relation, one can eliminate one term and get a smaller relation, leading to a contradiction (unless that smaller relation is trivial, in which case some desired condition is satisfied). You have doubtless seen this method used to prove that eigenvectors of a linear map corresponding to distinct eigenvalues are linearly independent. Actually, as noted in the comments to p.562 in this Companion, linear independence of eigenvectors and linear independence of characters are essentially the same result. But I sketch below another application of the above technique which I do not see how to regard as giving ''the same result''. In fact, it gives an alternative proof (also used in some texts) of the step in the development of the Fundamental Theorem of Galois Theory complementary to the one mentioned above; namely, that if k is the fixed field of a finite group G of automorphisms of a field E, then $[E:k] \leq (G:1)$ (cf. Lang, Theorem VI.1.8, p.264).

To get this, let $G = \{g_1, \dots, g_n\}$. Then we want to show that any $r > n$ elements $\alpha_1, \dots, \alpha_r \in E$ are k-linearly dependent. Consider the r n-tuples $(g_1(\alpha_1), \dots, g_n(\alpha_1))$, $\dots$, $(g_1(\alpha_r), \dots, g_n(\alpha_r)) \in E^n$. As $r > n$ elements of an n-dimensional E-vector space, these must be E-linearly dependent. Let $c_1, \dots, c_r \in E$ be coefficients in a nontrivial linear relation that they satisfy.

This condition on $c_1, \dots, c_r$ says that for every $g \in G$, $c_1 g(\alpha_1) + \dots + c_r g(\alpha_r) = 0$. Applying g^{-1} to this relation, and writing $g^{-1} = h$, we see that this is equivalent to saying that for every $h \in G$, $h(c_1)\alpha_1 + \dots + h(c_r)\alpha_r = 0$. Clearly, the set of r-tuples $(c_1, \dots, c_r)$ satisfying this condition is an E-subspace of E^r closed under the action of G.

Let us take a nonzero element of this subspace which minimizes the number of nonzero entries, normalized to have some entry equal to 1. By applying any member of G, and subtracting the result from the given element, we get an element with fewer nonzero entries; hence by the minimality assumption, this must be zero. This means that the given element is invariant under the action of G, i.e., that it has entries in k. Hence it gives the coefficients of a linear dependence relation among $\alpha_1, \dots, \alpha_r$ over k, as desired.

Another result that can be gotten by the same approach is that of Exercise XIII:L32, p.550 (though Lang's hint leads to a different proof, using the preceding exercise on that page. For still another method, *using* linear independence of characters, see VI.5:1.)

On characters into division rings (sketch, optional).

If G is a monoid and E a division ring, we can again regard the set of maps $G \to E$ as a vector space E^G over E (let us say a *left* vector space, since there are two choices); and we may again ask whether the subset consisting of the *characters* $G \to E$ is E-linearly independent.

Let us look at a simple case. In the division ring $\mathbf{H}$ of quaternions (described in the last comment to §V.1, above), the elements ± 1, $\pm i$, $\pm j$, $\pm k$ are fourth roots of unity (all but the first pair being primitive fourth roots), hence each of these elements determines a character $Z_4 \to \mathbf{H}$, sending the generator [1] of Z_4 to the chosen element. This gives us *eight* characters in the *four*-dimensional $\mathbf{H}$-vector-space $\mathbf{H}^{Z_4}$. Hence these characters must be linearly dependent! (There are in fact more than eight such characters: an easy calculation shows that a quarternion $a + bi + cj + dk$ is a primitive fourth root of unity if and only if $a = 0$ and $b^2 + c^2 + d^2 = 1$, so the 4-dimensional space $\mathbf{H}^{Z_4}$ contains *uncountably* many characters.)

Surprisingly, however, there is a version of linear independence of characters that works in division rings. In the above example, one can show that all the primitive fourth roots of unity in $\mathbf{H}$ are *conjugate* to one another. (For example, note that $(1+i)j = k(1+i)$, so $j = (1+i)^{-1}k(1+i)$.) And conjugacy turns

out to be the key idea. Since the set of functions f from a monoid G to a division ring E can be multiplied on the right as well as on the left by elements of E, one can conjugate such a function f by an element $d \in E$, getting a function $x \mapsto d^{-1}f(x)d$, and it is easy to see that if f is a character, so are all its conjugates. Now if $f_1, \dots, f_n$ are characters $G \to E$, and we take any left linear relation $a_1 f_1 + \dots + a_n f_n = 0$ $(a_i \in E)$ that they satisfy, and collect terms according to the conjugacy classes of the characters f_i, one can prove that the sum over *each* conjugacy class will be zero; that is, every case of linear dependence among characters results from linear dependence *within* conjugacy classes. It remains to determine when a family of mutually conjugate characters, $d_1^{-1}fd_1$, ..., $d_n^{-1}fd_n$, is linearly dependent. This turns out to happen if and only if the elements $d_1, \dots, d_n \in E$ are left linearly dependent over the *centralizer* C in E of the submonoid $f(G)$. Indeed, given a C-linear dependence relation among the d's, you should not find it hard to write down an E-linear dependence relation among the above mutually conjugate characters. The proof of the converse result, and of the fact that linear dependence relations among arbitrary families of characters all come from such relations among families of conjugate characters, are obtained by a fairly natural adaptation of the argument given by Lang in the commutative case; you might try to discover the details yourself. Note that when E is a field, every conjugacy class consists of a single character, so in this case the above assertion that all linear dependence relations are consequences of such relations within conjugacy classes is equivalent to the linear independence of characters.

Re §VI.5. The norm and trace.

P.284, beginning of §VI.5 [<]: Suppose α is an element of a finite separable extension E of a field k, and we take an algebraic closure k^{a} of k containing E, and look at the set of distinct *conjugates* of α in k^{a}; i.e., the images of α under all embeddings $E \to k^{\mathrm{a}}$. Then we see that both the *sum* and the *product* of these conjugates are invariant under $G(k^{\mathrm{a}}/k)$; hence both lie in the base-field k.

These two elements are useful in studying α, but there are problems when we consider more than one element at a time. For instance, we would like to be able to say that for $\alpha, \beta \in E$, the sum of the conjugates of $\alpha+\beta$ equals the sum of the conjugates of α plus the sum of the conjugates of β. But if α, β and $\alpha+\beta$ do not all have the same stabilizer in the Galois group, it may not be possible to pair off ''corresponding'' conjugates of these elements, as we would need to do to prove equality; and, indeed, the desired equality will not in general hold. What will work, however, is to associate to each $\alpha \in E$ the sum, respectively the product, of $\sigma(\alpha)$ as σ ranges over all embeddings $E \to k^{\mathrm{a}}$. These may be thought of as the sum and product of the conjugates of α ''counting multiplicities''. The multiplicities are such that we always have a sum or product of $[E:k]$ terms, regardless of the degree of the particular element α. This sum and product are called the *trace* and *norm* of α with respect to the extension E/k.

If we want to extend these concepts to possibly inseparable extensions, there are two plausible ways to go: We could again take the sum and product of $\sigma(\alpha)$ over all embeddings $\sigma\colon E \to k^{\mathrm{a}}$, or we could take the sum and product of all conjugates of α using a common multiplicity that makes their total number come to $[E:k]$. These give different values, because an inseparable extension E of k has fewer than $[E:k]$ distinct embeddings in k^{a}. Of these two approaches, the second turns out to have the better algebraic properties (the first may not even give elements of k), though the resulting functions are not as useful in the inseparable as in the separable case. (In particular, the trace is identically zero when E is inseparable over k.)

Thus, when you read the definitions of these functions in §VI.5, you should recognize the terms $p^\mu = [E:k]_i$ as supplying the extra multiplicity needed in the inseparable case. (You might find it easiest to understand this section if you first work through it assuming E/k is separable, so that $p^\mu = 1$, then reread it with the general case in mind.)

P.284, last two displays [~]: Lang writes (starting with the 4th printing) $N_{E/k}(\alpha) = N_k^E(\alpha) = \dots$, and a similar formula for the trace, meaning that he will sometimes write these one way, and sometimes the other. (In earlier printings, he only had the second definition, but still used both notations.)

There is a general convention of mathematical typesetting, that one-letter symbols, when not in a special font such as boldface or script, are set in italics, while a symbol of more than one letter, if not in a special font, is set in roman. This is the reason the symbol N for ''norm'' is italic, while the symbol Tr for ''trace'' is roman. (Other examples of multi-letter symbols set in roman, used in Lang or elsewhere, are Irr, Hom, det, lim, sin, and log.)

The value of this convention is that if one sees an italic multi-letter expression such as Re, one can recognize it as a combination (e.g., denoting the left ideal of the ring R generated by an element e) rather than a single symbol (e.g., ''Re'' might denote ''the real part of –''). (Some French and German mathematical works use a variant convention, in which symbols consisting of a single *capital* letter are also set in roman, so that only single lower-case symbols are italicized; thus, in such works the above left ideal would be written Re. There are still other conventions used by particular authors – for instance, some authors feel one should use roman letters for symbols with standard meanings, so that they write the base of the natural logarithms as e, and the derivative of y as dy/dx; or that one should interchange the roles of roman and italic symbols in italicized passages, so that what is ''sin x'' in ordinary text becomes ''*sin* x'' in the statement of a theorem. But I don't recommend either of those conventions.)

P.286, proof of Theorem VI.5.2 [>]: Here Lang has in effect repeated the proof of Theorem III.6.4.

After that proof, read the three-line **Remark** on p.287, then come back to Corollary VI.5.3 and its proof (which in effect repeats the proof of Theorem III.6.1), then skip Corollary VI.5.4, which, as we noted earlier, belongs in the preceding section, and go on to Proposition VI.5.5.

P.287, proof of Proposition VI.5.5, first display [=]: Lang is very terse about the derivation of this formula. It can be looked at as an example of *Lagrange interpolation*. Recall that that is a technique for finding a polynomial of degree $< n$ that assumes n specified values at n specified points. One does this by first writing down n polynomials, each of which is zero at $n-1$ of the points and nonzero at the remaining one, then taking the unique linear combination of these that gives the right values at all the points. In this case, the n points are $\alpha_1, \dots, \alpha_n$, and a polynomial which is zero at all of these except α_i is $f(X)/(X-\alpha_i)$. To get the correct linear combination, we need to know the values of these polynomials at the α_i. If we write $f(X)/(X-\alpha_i) = g_i(X)$, then by differentiating the equation $(X-\alpha_i)g_i(X) = f(X)$ with respect to X and substituting $X = \alpha_i$, we find that $g_i(\alpha_i) = f'(\alpha_i)$; now the desired formula is easily deduced.

P.287, last display [~]: Here Lang is writing i for what he has called r so far.

It is curious that this proposition has established a dual basis for the powers of α in the field $k(\alpha)$ by performing a computation in the larger field $k(\alpha_1, \dots, \alpha_n)$.

A particular consequence of this result is that none of the β_i are zero; in fact, that they are linearly independent over k. Can you see a quick proof of either of these facts not using the proposition?

P.288, end of top paragraph [>]: Suppose E/k is a finite extension, and α an element of E. As noted in the above paragraph of Lang, multiplication by α can be looked at as a k-linear map $m_\alpha : E \to E$. Since we shall soon be looking at α as a member of two different fields, let us denote this operator on E more precisely as $m_{\alpha, E}$. What is the relation between the *minimal polynomial* of α and the *characteristic polynomial* of $m_{\alpha, E}$? (Recall from undergraduate linear algebra that the latter is the determinant of the matrix $XI - M_{\alpha, E}$, where I is the identity matrix, and $M_{\alpha, E}$ is the matrix representing $m_{\alpha, E}$ with respect to any k-basis of E.)

Let us first consider the case $E = k(\alpha)$. In that case, a basis for E over k is given by $1, \alpha, \dots, \alpha^{d-1}$, where $d = [E : k]$. It is easy to write down the matrix of $m_{\alpha, E}$ in terms of this basis, since that map sends the basis element α^i to the basis element α^{i+1} for each $i < d-1$, while it sends α^{d-1} to α^d, which is not one of the basis elements, but can be written as a linear combination of these elements using coefficients which are negatives of the corresponding coefficients of $\mathrm{Irr}(\alpha, k, X)$. The resulting matrix $M_{\alpha, E}$ is called the *companion matrix* of $\mathrm{Irr}(\alpha, k, X)$, and it is easy to compute its characteristic polynomial, which turns out to be precisely $\mathrm{Irr}(\alpha, k, X)$.

(Another way to show that the characteristic polynomial of $m_{\alpha,E}$ is $\mathrm{Irr}(\alpha, k, X)$, without computation, is to call on a result of linear algebra which you may have seen as an undergraduate, the *Cayley-Hamilton Theorem*, Theorem XIV.3.1 in Lang. This says that if m is a linear transformation on a finite-dimensional vector space, and f its characteristic polynomial, then $f(m) = 0$. Now since the map sending each $\beta \in E$ to $m_{\beta,E}$ is a k-algebra homomorphism from E to the algebra of k-vector-space endomorphisms of E, and $\mathrm{Irr}(\alpha, k, X)$ generates the ideal of polynomials satisfied by α in E, it also generates the ideal of polynomials satisfied by $m_{\alpha,E}$ in that endomorphism ring. But by the Cayley-Hamilton theorem, $m_{\alpha,E}$ satisfies its characteristic polynomial, hence the latter polynomial must be a multiple of $\mathrm{Irr}(\alpha, k, X)$. Since both these polynomials are monic of degree $[k(\alpha):k]$, they are equal, as claimed.)

In the case where E is not necessarily generated by α, let $w_1, \dots, w_r$ be any basis of E over $k(\alpha)$, and let us take as a basis of E over k the set of products of the w's with the elements of some basis for $k(\alpha)$ over k. We find that the matrix representing $m_{\alpha,E}$ in terms of such a basis consists of a diagonal array of r copies of the matrix of $m_{\alpha,k(\alpha)}$. The characteristic polynomial of such a diagonal array is the product of the characteristic polynomials of the diagonal blocks, hence we find that in this general case, the characteristic polynomial of $m_{\alpha,E}$ is $\mathrm{Irr}(\alpha, k, X)^r$. i.e., $\mathrm{Irr}(\alpha, k, X)^{[E\,:\,k(\alpha)]}$.

Now it is easy to check that the linear map $m_{\alpha,k(\alpha)}$ has trace equal to the negative of the term-after-the-leading-term of $\mathrm{Irr}(\alpha, k, X)^{[E\,:\,k(\alpha)]}$. Indeed, of the diagonal entries of the companion matrix, only one is not automatically zero, and that one has the desired value. Similarly, the determinant of that map is $(-1)^{[k(\alpha):k]}$ times the constant term of $\mathrm{Irr}(\alpha, k, X)^{[E\,:\,k(\alpha)]}$. (Key to checking this: multiply the companion matrix by a matrix P that permutes the $[k(\alpha):k]$ rows cyclically so that the string of 1's moves onto the main diagonal. The determinant of the resulting matrix is easy to evaluate; note also that the matrix P we have multiplied by has determinant $(-1)^{[k(\alpha):k]+1}$.) But these coefficients, with the indicated adjustments of sign, are precisely $\mathrm{Tr}_{k(\alpha)/k}(\alpha)$ and $N_{k(\alpha)/k}(\alpha)$, by the definition of those trace and norm. Moreover, from the preceding paragraph we see that $m_{\alpha,E}$ has trace equal to $[E : k(\alpha)]$ times the trace of $m_{\alpha,k(\alpha)}$, and norm the $[E : k(\alpha)]$th power of the norm of $m_{\alpha,k(\alpha)}$. Now when $\mathrm{Irr}(\alpha, k, X)$ is raised to the power $[E : k(\alpha)]$, its term-after-the-leading-term and its constant term also get multiplied by $[E : k(\alpha)]$, respectively raised to that power. It follows that the statements that the trace and norm of the element α are the trace and determinant of the operation of multiplication by α, which we established for those operations on $k(\alpha)/k$, are also true on the arbitrary extension E/k. This is Proposition VI.5.6; but I thought it preferable that you see that these results on the trace and the norm are particular cases of a general observation relating the minimal and characteristic polynomials.

(I have often wondered whether one could get further interesting results by studying the other coefficients of the characteristic polynomial of $m_{\alpha,F}$.)

Incidentally, for a general (not necessarily irreducible) polynomial f, one defines its companion matrix in terms of the coefficients of f in exactly the same way. To interpret this matrix, note that the ring $k[X]/(f)$, even if it is not a field, still has a vector space basis $1, \alpha, \dots, \alpha^{\deg(f)-1}$, where α is the image of X; so the companion matrix still represents the action of multiplication by α with respect to that basis.

Re §VI.6. Cyclic extensions.

P.288, statement of Theorem VI.6.1 (Hilbert's ''Theorem 90'') [>]: Let me give you

A brief motivation for Hilbert's Theorem 90 and its proof.

We know that the norm function of a field extension is invariant under the Galois group: $N(\sigma(\alpha)) = N(\alpha)$, and is multiplicative: $N(\alpha\beta) = N(\alpha)N(\beta)$. Hence if we have an extension with cyclic Galois group, $G(K/k) = \langle\sigma\rangle$, we see that we can get elements of norm 1 by taking quotients $\alpha/\sigma\alpha$ for $\alpha \in K - \{0\}$. Can we think of any other way to get such elements? Well, $\gamma/\sigma^m\gamma$ will similarly have norm 1 for any $m \in \mathbf{Z}$, $\gamma \in K - \{0\}$. But if we set $\alpha = \gamma(\sigma\gamma) \dots (\sigma^{m-1}\gamma)$, we see that $\gamma/\sigma^m\gamma = \alpha/\sigma\alpha$, so

this is still an example of the sort we first noted. Thus it seems plausible that expressions of the form $\alpha/\sigma\alpha$ give all elements of norm 1. Theorem VI.6.1 establishes this result. (The result first appeared as ''Theorem 90'' in Hilbert's book *Bericht über die Theorie der algebraischen Zahlkörper*, and has come to be called ''Hilbert's Theorem 90''.)

The proof Lang will give may be motivated as follows. Given β with norm 1, we want an α such that $\beta = \alpha/\sigma(\alpha)$. We can translate this to say we want to find an element α fixed by the k-linear map $\alpha \mapsto \beta\sigma(\alpha)$. Denoting this k-linear map φ, a quick computation shows that for all α, $\varphi^n(\alpha) = N(\beta)\,\sigma^n(\alpha)$, which by our hypotheses is α. Thus $0 = \varphi^n - 1 = (\varphi-1)(\varphi^{n-1} + \ldots + \varphi + 1)$. The second factor in this product is nonzero, by the linear independence of characters, hence has a nonzero element α in its image; this will lie in the kernel of $\varphi - 1$, i.e., will be an element of the form we want. You can verify that this operator $\varphi^{n-1} + \ldots + \varphi + 1$ is in fact the map used in the proof in Lang (first display on p.289).

P.289, end of proof of Theorem VI.6.1 [>]: Let us note

Another motivation for Hilbert's Theorem 90.

Hilbert's Theorem 90 can also be looked at in terms of a goal we are moving toward, and will achieve in the next section. This is to find necessary and sufficient conditions on a finite algebraic extension E/k for the elements of E to be *expressible in radicals*; i.e., for E to lie within a finite extension which is the top of a tower of fields, having k at the bottom, and such that each field in the tower is obtained from the field just below it by adjoining an element which has some power in that lower field. Now the properties of adjoining to a field k an nth root α of an element $a \in k$ (where n is not divisible by the characteristic of k) are simplest if our base field contains a primitive nth root of unity ζ_n. In this case, the extension is automatically Galois, since $X^n - a$ splits as $\prod (X - \zeta_n^i \alpha)$. I claim that if a is not an mth power in k for any $m > 1$ dividing n, then the Galois group is generated by an automorphism σ carrying α to $\zeta_n \alpha$. Indeed, the set of nth roots of unity (powers of ζ_n) such that our extension has an automorphism carrying α to itself times that root of unity is clearly a subgroup of $\langle \zeta_n \rangle$; thus it will be generated by ζ_n^m for some divisor m of n. We see that $\alpha^{n/m}$ is invariant under the Galois group, hence lies in k, and a is the mth power of this element, hence by our assumption on a, $m = 1$; this says that the Galois group has an element σ of the desired form. Further, each element of the Galois group is determined by which element $\zeta_n^i \alpha$ it sends α to, and all possibilities are achieved by powers of our σ, hence the Galois group is precisely $\langle \sigma \rangle$.

What about the converse? If k is a field containing a primitive nth root of unity ζ_n, and E a Galois extension of k with cyclic Galois group $\langle \sigma \rangle$ of order n, is E generated over k by an nth root of an element of k? It is not hard to see that this is equivalent to asking whether E has a nonzero element α such that $\sigma(\alpha) = \zeta_n \alpha$. Note that such a relation at least passes the test of consistency, in that applying σ successively n times would bring α back to α.

Rather than looking at this question in isolation, let us look generally at elements $\beta \in E$ (not necessarily in k) such that it would it be ''consistent'' for a nonzero element $\alpha \in k$ to satisfy

$$\sigma(\alpha) = \beta\alpha. \tag{c35}$$

To find the ''consistency'' condition needed, we compute

$$\begin{aligned} \sigma^0(\alpha) &= \alpha, \\ \sigma^1(\alpha) &= \beta\alpha, \\ \sigma^2(\alpha) &= \sigma(\beta)\beta\alpha, \\ &\ldots \\ \sigma^n(\alpha) &= \sigma^{n-1}(\beta) \ldots \sigma(\beta)\beta\alpha, \end{aligned}$$

Since $\sigma^n = 1$, the last equation simplifies to $\alpha = N(\beta)\alpha$, so the desired consistency condition is

$N(\beta) = 1$. Hilbert's Theorem 90 is the statement that this necessary condition is also sufficient for the existence of an element α satisfying (c35).

Returning to the particular case where $\beta = \zeta_n$ is a primitive nth root of unity in the base field k, we conclude that E will indeed contain an element α satisfying $\sigma\alpha = \zeta_n \alpha$, equivalently, an element whose nth power, but no lower power, satisfies $a \in k$; so $E = k(a^{1/n})$. The details are given in the proof of the next result in Lang, Theorem VI.6.2.

P.290, first sentence [~]:

Motivation for the additive form of Hilbert's Theorem 90.

We have seen that Hilbert's Theorem 90 is useful in studying cyclic Galois extensions where the base field contains a primitive nth root of unity. If we start with a cyclic extension K/k where k does not have such a root of unity, we may still be able to get information by adjoining such a root, and looking at what Lang calls the ''lifted'' extension $K(\zeta_n)/k(\zeta_n)$. But a situation where even this is not possible is if n is divisible by the characteristic of k, since no field of characteristic dividing n can contain a primitive nth root of unity.

What will a cyclic field extension look like in this case? The simplest situation to consider is that of an extension K/k of fields of characteristic p with Galois group $\langle\sigma\rangle$ cyclic of order p. Since we cannot find elements of K on which σ acts by multiplication by a primitive pth root of unity, what other plausible behavior can we come up with that, when applied p times, would bring us back to the original element?

Well, the behavior $\sigma(\alpha) = \alpha+1$ would have this property; so we may ask whether every cyclic extension of degree p in characteristic p has an element α satisfying this equation. Again, we can generalize: The condition on an element $\beta \in K$, where K/k is a cyclic extension of *any* order, and of *any* characteristic, that is needed for the equation $\sigma(\alpha) = \alpha+\beta$ to be consistent with $\sigma^n = 1$, is easily seen to be $\mathrm{Tr}(\beta) = 0$. The next result that Lang will prove, the ''additive form of Hilbert's Theorem 90'', says that this, too, is sufficient as well as necessary.

There is another way to look at that result, analogous to the ''brief motivation for Hilbert's Theorem 90'' that I gave in my comment on Lang's p.288. In that paragraph, if you replace ''norm'' by ''trace'', ''multiplicative'' by ''additive'', and ''quotient'' by ''difference'', you will get a brief motivation for Theorem VI.6.3.

P.290, first display [=]: As in the proof of the multiplicative form of Hilbert's Theorem 90, Lang uses exponential notation for the action of σ when writing out the formula for the desired element α; but this time he is inconsistent, mixing both notations in the same formula.

P.290, statement of Theorem VI.6.4 [<]: On the preceding page of Lang, when the multiplicative form of Hilbert's Theorem 90 was used to get an element α satisfying $\sigma(\alpha) = \zeta_n \alpha$, we noted that this would have minimal polynomial of the form $X^n - a$. If we similarly apply the additive form of Hilbert's Theorem 90 to get an element α satisfying $\sigma(\alpha) = \alpha+1$, what form will its minimal polynomial take? Well, the conjugates of α will be α, $\alpha+1$, ..., $\alpha+p-1$, hence that polynomial can be written $\prod_{i=0,\dots,p-1} (X-\alpha+i)$. To simplify this, let $Y = X-\alpha$; then it becomes $\prod (Y+i)$, i.e., the polynomial of degree p that has all elements of $\mathbf{Z}_p$ as roots. We recall that this polynomial is $Y^p - Y$. Hence $\prod (X-\alpha+i) = (X-\alpha)^p - (X-\alpha) = X^p - X - (\alpha^p - \alpha)$. This leads to the statement of Theorem VI.6.4.

P.290, next to last line [=]: ''$1, \dots, p$'' should be ''$1, \dots, p-1$''.

P.291, top paragraph [<]: The argument can be completed more briefly. After the first three words, continue: ''Indeed, σ must take α to $\alpha+i$ for some $i \in \{1, \dots, p-1\}$, and under addition, i generates the group $\{0, \dots, p-1\}$, hence $\langle\sigma\rangle$ is transitive on the set of roots $\{\alpha+i\}$ of f, proving irreducibility.''

P.291, end of §VI.6 [>]. Here are some further notes on the subject of this section.

Some exact sequences (optional).

Let K/k be a Galois extension with cyclic Galois group $\langle\sigma\rangle$. Hilbert's Theorem 90 can be looked at as saying that if we regard the norm function as a group homomorphism $K^* \to k^*$, then its kernel is the image of the endomorphism $\alpha \mapsto \sigma(\alpha)/\alpha$ of the group K^*. Now we know the kernel of that endomorphism; it is the set of elements $\alpha \in K^*$ fixed under σ, i.e., k^*. These results can be summarized by saying that the sequence of group homomorphisms

$$1 \to k^* \xrightarrow{\subseteq} K^* \xrightarrow{\alpha \mapsto \sigma(\alpha)/\alpha} K^* \xrightarrow{N} k^* \tag{c36}$$

is exact.

This suggests the question: what can one say about the *cokernel* of the rightmost arrow above?

One case where we can answer this easily is if k and K are finite fields. Note that in an exact sequence of finite groups, each group is an extension of the image of the map coming into it by the image of the map going out of it. From this one can deduce that in a finite exact sequence beginning and ending with trivial groups, the product of the orders of the groups in even-numbered positions equals the product of the orders of the groups in odd-numbered positions. Thus, if we let C denote the cokernel of the last map of (c36), and tack $\to C \to 1$ onto that sequence, we get $(k^* : 1)\cdot(K^* : 1)\cdot(C : 1) = (K^* : 1)\cdot(k^* : 1)$; hence $(C : 1) = 1$. So if our fields are finite, the last map of (c36) is surjective.

If the fields are infinite, that map may or may not be surjective. For a negative example, let $k = \mathbf{R}$, $K = \mathbf{C}$. Then since the range of the norm map on nonzero complex numbers consists of the *positive* real numbers, the cokernel of the right-hand map of (c36) is isomorphic to Z_2.

The observations we made about exact sequences of finite groups also apply to exact sequences of *finite-dimensional* vector spaces, with sums of *dimensions* replacing products of group-orders. Since both fields involved in the additive form of Hilbert's Theorem 90 are finite-dimensional over k, we can conclude that the sequence

$$0 \to k \xrightarrow{\subseteq} K \xrightarrow{\alpha \mapsto \sigma(\alpha) - \alpha} K \xrightarrow{\mathrm{Tr}} k \to 0 \tag{c37}$$

is always exact. (Note that we generally use ''1'' to denote the trivial group when groups are written multiplicatively, ''0'' when they are written additively.)

Some linear algebra.

The special cases of Hilbert's Theorem 90 used in proving Theorems VI.6.2 and VI.6.4, namely the cases where $\beta = \zeta_n$ in the multiplicative version, and where $\beta = 1$ in the additive version, can be looked at in yet another way. If K is a cyclic extension of k, with Galois group $\langle\sigma\rangle$ of order n, then $\sigma\colon K \to K$, regarded as a linear transformation on the k-vector space K, satisfies the polynomial $t^n - 1$. If k has a primitive nth root of unity, this polynomial splits into *distinct* linear factors over k, hence the k-vector-space K has a basis B of eigenvectors of σ. I claim that no two of these eigenvectors have the same eigenvalue. For if distinct basis elements $\alpha, \beta \in B$ satisfied $\sigma(\alpha) = \zeta_n^i \alpha$ and $\sigma(\beta) = \zeta_n^i \beta$ with the same i, then α/β would be an element not in k which was fixed under σ, contradicting the Fundamental Theorem of Galois Theory. In particular, the n eigenvalues must comprise all n roots of $t^n - 1$, so some eigenvector α will have eigenvalue ζ_n, as desired. (We can see that up to scalars in k, the other eigenvectors will be the powers of α.)

In the additive case, where k has characteristic p, the polynomial $t^p - 1$ satisfied by σ factors as $(t - 1)^p$. A linear transformation satisfying this polynomial has Jordan canonical form in which the nonzero entries consist of 1's down the main diagonal, and possibly some 1's in the next diagonal up. Since $\sigma \neq 1$, there must be at least one 1 above the main diagonal; looking at the ''highest'' of these, we see that the corresponding basis vector will be sent by σ to itself plus the preceding vector in our basis, which will be a fixed element. Thus, we get an α such that $\sigma(\alpha) = \alpha + c$ for some $c \in k$, and dividing by c, we get an element which adds 1 on applying σ, as in Theorem VI.6.4.

In §VI.10, some results are developed which can be regarded as analogs of Hilbert's Theorem 90 for noncyclic Galois extensions. Although there is not time to cover that section in 250A, I have put some notes on it in this Companion which bring out this relationship. That section and the notes to it may be read once one has read this section.

Re §VI.7. Solvable and radical extensions.

P.293, end of proof of Theorem VI.7.2 [>]: The result just proved answers the famous question of when a polynomial equation can be ''solved in radicals''! Galois Theory in fact arose out of the effort to answer this question. Note that if an equation has degree ≤ 4, its Galois group can be identified with a subgroup of S_4, a solvable group; so every polynomial equation of degree ≤ 4 is solvable in radicals. On the other hand, S_5 is not solvable (Theorem I.5.5), and there exist polynomials of degree 5 over $\mathbf{Q}$ whose Galois group is this group (Example 4 on p.272; Example 6 on p.273). Hence polynomials of degree 5 (and higher) are not in general solvable in radicals.

Another classical question which can be answered by Galois Theory is, ''Which lengths can be constructed from a given set of lengths by ruler and compass; and in particular for which n can one construct by ruler and compass a regular n-gon?'' For developments of these results, see, e.g., Hungerford's *Algebra*, or Stewart's *Galois Theory*.

Re §VI.8. Abelian Kummer theory.

This subject is named for Ernst Kummer (1810-1893).

P.293, first sentence of the section [=]: The phrase ''the theorem concerning cyclic extensions'' refers to Theorem VI.6.2 and its characteristic-p analog, Theorem VI.6.4.

P.294, 4th paragraph [~]: Note that if $a' = b^m a$ $(a, a', b \in k - \{0\})$, then to adjoin to k an mth root of a' is equivalent to adjoining an mth root of a. Thus, in classifying fields obtained by adjoining mth roots, we should use some system of elements which contains $b^m a$ if and only if it contains a. This is the idea behind looking at subgroups $B \subseteq k - \{0\}$ which contain all mth powers.

P.294, 5th paragraph [=]: In developing the properties of the map $\sigma \mapsto \omega_\sigma$, Lang fixes α and does not show it in the symbol ω_σ, though it appears in the definition thereof. Later in the paragraph, he switches to considering the dependence of this root of unity on both σ and α. He notes that this does not change if α is replaced by a different mth root of a, hence it can be regarded as a function of σ and a, which he writes $\langle \sigma, a \rangle$. Intuitively, this definition says $\langle \sigma, a \rangle = \sigma(a^{1/m})/a^{1/m} \in \boldsymbol{\mu}_m$, but because of the ambiguity of the symbol $a^{1/m}$, the actual definition must be a bit wordier. This operation turns out to be a bilinear map into a cyclic group of order m, so the duality of §I.9 becomes applicable; however, note that the groups here are written multiplicatively, in contrast to the notation of that section.

P.295, first display [=]: For ''$B/k^{*m} \approx G\hat{}$'' read ''$G \approx (B/k^{*m})\hat{}$''. Both are true, but the latter is the isomorphism he proves, and also the one that goes over to the infinite case.

P.295, fourth line of proof of Theorem VI.8.2 [=]: Lang says, ''Thus we may assume without loss of generality that B_2/k^{*m} is finitely generated, hence finite''. What he is asserting is that if we knew the desired result ''$k(B_1^{1/m}) \subseteq k(B_2^{1/m}) \Rightarrow B_1 \subseteq B_2$'' in the case where B_2/k^{*m} is finite, then by using the observation in the preceding sentence, we could deduce the same result in the general case. (Do you see how?) Since to prove that result it suffices to prove it under the assumption that B_2/k^{*m} is finite, that assumption will be made for the remainder of the paragraph.

Later in the proof, at the first display, change ''or'' to ''and also''.

P.295, sentence beginning in third line of last paragraph, ''Any finite subextension...'' [=]: The reasoning here is: Any finite subextension of K will still be Galois (Corollary VI.1.11), and its Galois group, being a finite homomorphic image of $\mathrm{Gal}(K/k)$, will be a finite abelian group of exponent p. Such a group is a direct product of cyclic abelian groups of exponent p (Theorem I.8.2), so by Corollary VI.1.6, our subextension will be a compositum of subfields with these Galois groups.

P.296, first display [~]: Why are we so interested in the operation $x \mapsto x^p - x$ that we invent a symbol for it? Because Theorem VI.6.4 showed us that every cyclic extension of degree p in characteristic p is

generated by an element x satisfying $x^p - x = a$ for some a in the base field. Thus, the operation $x \mapsto x^p - x$ in characteristic p is analogous to the pth power operation in characteristic $\neq p$.

P.296, Theorem VI.8.3, definition of $\langle \sigma, a \rangle$ [~]: Just as, in the characteristic-not-dividing-m case, we made a definition which intuitively said $\langle \sigma, a \rangle = \sigma(a^{1/m})/a^{1/m} \in \boldsymbol{\mu}_m$, so here the idea is $\langle \sigma, a \rangle = \sigma(\wp^{-1}a) - (\wp^{-1}a) \in \mathbf{Z}_p$. Again this map turns out to be bilinear, so that we will be able to apply the duality of §I.9. (And this time, the groups in question are written additively.)

P.296, last paragraph [>]: In the characteristic-not-dividing-m case, there was no extra difficulty when m was divisible by the square of a prime, since one can still have primitive mth roots of unity for such m. But in characteristic p, while the map $\alpha \mapsto \alpha+1$ has order p, there is no obvious similar construction of order p^2. Hence more elaborate machinery is needed, for which Lang refers us to the exercises. (I have not had a chance to go through those and see how workable they are.)

P.296, end of §VI.8 [>]:

Overview of the above section, and one further result.

Suppose k is a field containing a primitive mth root of unity, and one adjoins to k the mth roots of a family of elements of k. Then the behavior of any element σ of the resulting Galois group is determined by specifying which mth roots of unity each of these adjoined elements gets multiplied by. Obviously, there are certain consistency conditions that must be satisfied. For instance, in $\mathbf{Q}(2^{1/2}, 3^{1/2}, 6^{1/2})$, what an automorphism σ does to $6^{1/2}$ will be determined by what it does to $2^{1/2}$ and $3^{1/2}$. Generally speaking, we can say that if we have adjoined mth roots $\alpha_1, \dots, \alpha_r$ of elements $a_1, \dots, a_r \in k$ (where this list may involve repetitions), and if the product $a_1 \dots a_r$ is already an mth power in k, then $\alpha_1 \dots \alpha_r$ must lie in k; hence the roots of unity by which σ multiplies these factors must have product 1.

Kummer theory tells us that these consistency conditions are the only restrictions on such automorphisms. To express this, one lets B denote the multiplicative group generated by all elements whose mth roots one wishes to adjoin to k, together with all nonzero elements which already have mth roots in k. For each element σ of the Galois group, one defines a homomorphism $\langle \sigma, - \rangle : B \to \boldsymbol{\mu}_m$, sending each $a \in B$ to the root of unity by which σ multiplies the mth roots of a. Theorem VI.8.1 says that, letting G denote $G(k(B^{1/m})/k)$, the resulting bilinear map

$$\langle -, - \rangle : \ G \times B \ \to \ \boldsymbol{\mu}_m \tag{c38}$$

has kernels 1 and k^{*m} respectively on the left and right. Hence, if B/k^{*m} is finitely generated, equivalently, finite, we can conclude by Theorem I.9.2 that (c38) induces an isomorphism between G and the dual group to B/k^{*m} with respect to $\boldsymbol{\mu}_m$. Elements of this dual group represent possible ''consistent'' behaviors for maps $\langle \sigma, - \rangle$; hence every such behavior is realized by some σ.

Recall that the relation between subextensions of a Galois field extension and subgroups of its Galois group is order-*reversing*. The relation between subgroups of a finite abelian group and subgroups of its dual group is also order-reversing; hence Kummer theory yields an order-*preserving* correspondence between subextension of $k(B^{1/m})$ and subgroups of B containing k^{*m}.

What if B/k^{*m} is not finitely generated? In general, when we have a bilinear map from two abelian groups into a cyclic group, and we pass to the factor-groups of these groups by the appropriate kernels of the bilinear map, then each factor-group can be identified with a *subgroup* of the dual of the other. But I claim that in the present situation, we get an identification of the Galois group G with the whole of $\mathrm{Hom}(B/k^{*m}, \boldsymbol{\mu}_m)$. I will sketch the argument, which is another example of a ''compactness proof'' (cf. comments in this Companion on p.235, end of §V.2 of Lang), i.e., a proof that a certain infinite set of conditions will have a model if all its finite subsets do.

Given $h \in \mathrm{Hom}(B/k^{*m}, \boldsymbol{\mu}_m)$, we want to find a $\sigma \in G$ which induces h. Now for every subgroup $B_0 \subseteq B$ finitely generated over k^{*m}, the restriction of h to B_0 is a member of $\mathrm{Hom}(B_0/k^{*m}, \boldsymbol{\mu}_m)$, and so by the results proved in Lang, there is a unique $\sigma_{B_0} \in G_{B_0} = G(k(B_0{}^{1/m})/k)$ which induces this map. We see that given B_0 and B_1 each finitely generated over k^{*m}, the automorphisms σ_{B_0} and

σ_{B_1} agree on $k((B_0 \cap B_1)^{1/m})$. From the fact that B is the union of all its subgroups B_0 finitely generated over k^{*m}, it is not hard to deduce that these automorphisms fit together to give a single automorphism of $k(B^{1/m})$, which induces h, as claimed. (We ''put them together'' by taking their union as sets.)

(The analogs of the above considerations likewise apply to the case of exponent-p abelian extensions of a field of characteristic p.)

Re §VI.9. The equation $X^n - a = 0$.

P.297, statement of Theorem VI.9.1 [~]: This theorem gives necessary and sufficient conditions for a polynomial of the form $X^n - a$ to be irreducible. To motivate it, let us consider how such a polynomial might factor.

Clearly, if $X^n - a$ is reducible in $k[X]$, then its irreducible factors all have degrees $< n$, and conversely. (Incidentally, not all these factors need have the same degree, an extreme case being when $a = 1$.) So what we are asking is how an nth root (in an extension field) of an element $a \in k$ can have degree $< n$ over k. An obvious case where this happens is when $a = b^m$ where $b \in k$, $m \mid n$, and $m > 1$; then we can get a root of $X^n - a$ of degree $< n$ by taking an n/mth root of b. To say a is an mth power for some $m \mid n$ with $m > 1$ is equivalent to saying it is a pth power for some prime $p \mid n$, and this is the formulation used in the theorem.

There is another less obvious way $X^n - a$ can be reducible. It arises from the fact that $(1+i)^4 = -4$, so that the polynomial $X^4 - (-4)$ has the root $1+i$, which only has degree 2 over $\mathbf{Q}$, though -4 is not a square in $\mathbf{Q}$. The resulting factorization of $X^4 - (-4)$ over $\mathbf{Q}$ can by Gauss's Lemma be taken to be a factorization over $\mathbf{Z}$ (the factors are $(X^2 - 2X + 2)(X^2 + 2X + 2)$, cf. Remark 2 on p.183), and since $\mathbf{Z}$ can be mapped homomorphically into any field, this gives a factorization of $X^4 - (-4)$ over any field k. This in turn gives a factorization of $X^4 - (-4)b^4$ for any $b \in k$, and hence of $X^{4m} - (-4)b^4$ for any m. Thus, another necessary condition for $X^n - a$ to be irreducible is that if n has the form $4m$, a should not have the form $-4b^4$. Theorem VI.9.1 asserts that the two necessary conditions just noted are also sufficient for $X^n - a$ to be irreducible.

P.297, first two paragraphs of the proof of Theorem VI.9.1 [~]: This argument, showing that the statement of the theorem is true *if* it is true whenever n is a prime power, can be replaced by the following simpler argument, pointed out to me by Lenstra. Let α be a root of $X^n - a$, and let the prime factorization of n be $p_1^{r_1} \ldots p_t^{r_t}$. For each i, $\alpha^{n/p_i^{r_i}}$ is a $p_i^{r_i}$th root of a lying in $k(\alpha)$. Assuming the theorem is true for prime power exponents, this element has degree $p_i^{r_i}$, hence $p_i^{r_i} \mid [k(\alpha):k]$. Since this is true for all i, we get $n \mid [k(\alpha):k]$. But $[k(\alpha):k] \leq n$, so $[k(\alpha):k] = n$.

After this argument, continue from ''We now suppose that $n = p^r$ ...'' (11 lines from the bottom). However, here is a note about the next paragraph:

P.297, third from last paragraph [~]: Here Lang considers the purely inseparable case. The ''even more trivial argument'' that he refers to uses the Frobenius map. (Can you see how to do this?) To deduce the final sentence, note that the preceding sentence gives $[k(a^{1/p^r}) : k(\alpha)] = p^{r-1}$. It follows that $[k(a^{1/p^r}) : k] = p^r$.

Lang's induction also omits the base case, $r = 1$. To see that case, note that if $\alpha = a^{1/p} \notin k$, then the irreducible polynomial of α, being a divisor of $X^p - a$ of degree > 1, must have multiple roots. Then Proposition V.6.1 tells us that the degree is divisible by p; so it is not a proper divisor of $X^p - a$.

P.297, last paragraph [=]: The first word, ''Suppose'', should be ''By assumption''. On the second line, ''a root'' should be ''the root''. On the third line, ''and $\alpha^p = a$. Taking the norm from $k(\alpha)$ to k'' should be ''. Taking the norm from $k(\alpha)$ to k of the equation $\alpha^p = a$''. On the last line of the page, the justification for ''it follows'' is: the set of integers i such that a^i is a pth power in k clearly forms an additive group. It contains p by definition, and the preceding line shows that it contains d. As

these integers are relatively prime, it contains 1.

P.298, second paragraph (''Assume ...'') [~]: Although one of our hypotheses is that a is not a pth power in k, this does not say whether or not, after we adjoin a pth root α of a to k, this α will be a pth power in the resulting field $k(\alpha)$, and the question becomes crucial to the proof. Hence Lang considers two cases. First he assumes it does not, and quickly completes the inductive argument. (Note that if p is odd, the assumption that α is not a pth power gives him the hypothesis for an inductive application of the theorem, while if $p = 2$, he must in addition show that α is not of the form $-4\beta^4$ in this field. I would simplify the argument for the latter case as follows: After ''If $p = 2$'' continue, ''... then our hypotheses says that α is not a square in $k(\alpha)$; hence applying the nontrivial automorphism of this field, we see that $-\alpha$ is also not a square, and so in particular cannot be written $4\beta^4$. Hence α cannot be written $-4\beta^4$ in $k(\alpha)$, and again by induction ...''.) The remainder of the proof of the theorem, starting with the last sentence of that paragraph, is devoted to showing that the case where α is a pth power in $k(\alpha)$ cannot occur.

P.298, before third-from-last display, ''In either case ...'' [=]: This is because -4 is a square in $k(i)$. The symbol $\pm$ is much used in the remainder of the proof. In each occurrence, it will simply mean ''either + or −'', with no implicit ''respectively''. That is, the ''+'' case of one occurrence need not correspond to the ''+'' case of another.

P.298, next-to-last display [>]: This contradiction has shown that the two factors of $X^{2^r} - a$ in $k(i)[X]$ are each irreducible in that ring, so by unique factorization in this ring, these are the only proper monic nonunit factors of $X^{2^a} - \alpha$ in $k(i)[X]$. Since these factors do not lie in $k[X]$, we conclude that the polynomial is irreducible in the latter ring, as desired.

P.299, end of line 2 [>]: To strengthen this result, after ''*or if p is odd*'' add a third case, ''*or if -1 is a square in k*''. (To see that the conclusion holds in this case, note that by the preceding cases, this case need only be considered when $p = 2$ and 2 is not the characteristic. In that case, the condition ''*a is not a p-th power*'' together with -1 being a square implies $a \notin -4k^4$, as required to apply the theorem.)

P.299, statement of Corollary VI.9.3 [<]: Theorem VI.9.1, whose proof has taken up most of this section, is moderately interesting; but the real reason we took the time to prove it is for the sake of Corollary VI.9.3, a surprising result saying that any field k such that $1 < [k^{\mathrm{a}} : k] < \infty$ must look very much like the single example that we know: the real numbers! Incidentally, that result is due to the same two people as Theorem VI.6.4, and the name ''the Artin-Schreier Theorem'' is applied to both statements.

P.299, second line of proof of Corollary VI.9.3 [=]: By ''inseparable over some subfield of degree > 1'' Lang means ''inseparable of degree > 1 over some subfield''.

P.299, middle of second paragraph of proof of Corollary VI.9.3 [>]: Lang refers to Exercise VI:L29; you should look over that exercise, and think about how to do it.

P.299, last four lines of second paragraph of proof of Corollary VI.9.3 [<]: This argument can be shortened with the help of the strengthened version of Corollary VI.9.2 mentioned above: Since k^{a} is the splitting field of $X^p - a$, and $i \in k_1 \subseteq F$, that corollary shows that $X^{p^2} - a$ is irreducible, contradicting the fact that $[k^{\mathrm{a}} : F] = p$.

P.299, third paragraph of proof of Corollary VI.9.3 [<]: Let me give the content of this paragraph in a way that yields more explicit information about k. First note that for any $a \in k - \{0\}$, either a is a square, or $-a$ is a square, but not both. (Indeed, let σ be the nontrivial automorphism of $k^{\mathrm{a}} = k(i)$. If we take a square root of a and a square root of $-a$ in this field, then each is either fixed, or sent to its negative by σ. But their product is $\pm ia$, which is sent to its negative by σ, so *one* of these elements must be fixed, and the *other* sent to its negative. Thus, one lies in k and the other does not.) Let us call $a \in k - \{0\}$ ''positive'' if a is a square, and ''negative'' if $-a$ is a square. We claim that a sum of two positive elements, say u^2 and v^2, is positive. Indeed, $u^2 + v^2$ can be written $(u + iv)(u - iv)$ in $k(i)$; note that neither of the factors is zero, so this element is nonzero. Since $k(i)$ is algebraically closed, we

can take a square root of $u+iv$; say $u+iv = (x+iy)^2$ $(x,y\in k)$. Applying σ, we have $u - iv = (x - iy)^2$; multiplying these equations together, we see that $u^2+v^2 = (x^2+y^2)^2$, which is the desired result. Since 1 is positive, it now follows by induction that the sum of any number of 1's is positive, hence nonzero; hence k has characteristic 0, completing the proof of Corollary VI.9.3.

In fact, if for $a, b\in k$, we let ''$a > b$'' mean that $a - b$ is positive, then it is easy to verify that this is a total ordering on k, which obeys the usual laws relating inequalities and arithmetic operations. A field k with such an ordering is called an ''ordered field''; these are considered in Chapter XI of Lang. The above paragraph is essentially the proof of Proposition XI.2.4 (p.452).

Corollary VI.9.3 shows that a field k which has ''few'' algebraic extensions, in the sense that $[k^{\mathrm{a}}:k]$ is finite, must have very restrictive properties. We remark, however, if we weaken the sense of ''few extensions'', there are wider classes of examples. In particular, Exercises VI:L26 and VI:L27 (p.325) show two ways of getting fields with the property that every finite extension is cyclic, and it is not hard to show that such a field has at most one extension of each degree.

P.299, third from last line [=]: This result, that any nontrivial torsion element must have order 2, is true not only for $G(\mathbf{Q}^{\mathrm{a}}/\mathbf{Q})$ but for $G(k^{\mathrm{a}}/k)$ for any field k. This follows from Corollary VI.9.3: If σ is an automorphism of finite order $n>1$, then by Artin's Theorem, $[k^{\mathrm{a}}:(k^{\mathrm{a}})^{\langle\sigma\rangle}] = n$; but by the corollary referred to, this can only happen for $n = 2$. From that corollary we also see that if k has positive characteristic *or* contains a square root of -1, then $G(k^{\mathrm{a}}/k)$ has *no* nontrivial elements of finite order.

P.300, top [<]: We began this section by determining when a polynomial of the form $X^n - a$ was irreducible. It is natural to ask, in such a case, what its Galois group is. Lang introduces this subject as Example 2, and then obtains an answer, under certain assumptions, in Theorem VI.9.4. The idea is: first adjoin a primitive nth root of unity ζ_n, *then* adjoin an nth root α of a. The resulting Galois group will be an extension of its subgroup $G(k(\zeta_n,\alpha)/k(\zeta_n))$ by its factor-group $G(k(\zeta_n)/k)$. Under certain conditions, the first of these will be isomorphic to the additive group of $\mathbf{Z}/n\mathbf{Z}$, and the second to the multiplicative group of units of this ring, and the extension will be the natural semidirect product of these two groups. (You might look at the Galois group of the splitting field of $X^3 - 3$ over $\mathbf{Q}$ from this point of view.) Lang will describe this semidirect product as a group of matrices over the ring $\mathbf{Z}/n\mathbf{Z}$.

P.301, first two lines [=]: I don't understand the reasoning Lang gives after ''because''. The assertion that the commutator is the subgroup shown at the bottom of the preceding page is only true for n odd – you can easily check that for $n=2$, the commutator subgroup is instead the trivial subgroup – but Lang's words don't seem to use the assumption that n is odd. In any case, for n odd, consider the commutator of an element as on the bottom of the preceding page with the element as in the third display of that page having $b = 0$ and $d = 2$. (When n is odd, $2\in(\mathbf{Z}/n\mathbf{Z})^*$.)

P.301, 2nd through 8th lines after the diagram [=]: Where Lang says ''We apply the first part of the proof to $X^p - \beta$ over $k(\beta)$'', you should verify that the hypotheses are satisfied here, in particular that $[k(\beta,\boldsymbol{\mu}_p):k(\beta)] = \varphi(p)$. (Hint: $[k(\boldsymbol{\mu}_p):k]$ and $[k(\beta):k]$ are relatively prime.) In the next sentence, the point is that $k(\beta,\boldsymbol{\mu}_n)$ is an abelian subextension of $k(\beta)$, hence so is the left-hand side of the displayed equation; but by ''the first part of the proof, applied to $X^p - \beta$ over $k(\beta)$'', the intersection of $k(\alpha)$ with the maximal abelian subextension of the splitting field of $X^p - \beta$ over $k(\beta)$ is just $k(\beta)$.

Re §VI.10. Galois cohomology.

P.302, start of §VI.10 [<]:

Motivation of the ideas of this section.

Hilbert's Theorem 90, which we studied in §VI.6, concerned *cyclic* Galois extensions. Is an analogous result true in the noncyclic case?

The answer depends on how one tries to generalize the theorem. If one focuses on the element β, and regards the theorem as a way of constructing the most general element of norm 1, then the obvious analog in a noncyclic Galois extension would say that every element β of norm 1 can be written as a

product, over various $\sigma \in G$ and $\alpha \in K^*$, of the elements $\sigma(\alpha)/\alpha$. This statement turns out to be false. (Here is the idea of a counterexample, shown to me by H. Lenstra. Start with a quadratic extension E/k such that E contains no square root of -1, but has an element β with norm $N_k^E(\beta) = -1$; the latter condition can be achieved by letting E be generated over k by a root β of an irreducible polynomial $X^2 + aX - 1$. Now take another quadratic extension F of k, and take their compositum EF within a common extension. Then $N_k^{EF}(\beta) = (-1)^2 = 1$; but if E and F are chosen properly, one can prove from the fact that $N_k^E(\beta) \neq 1$ that β is not a product of elements $\sigma(\alpha)/\alpha$ in the Galois extension EF.)

On the other hand, if we focus on the α in Hilbert's Theorem 90, we can look at that theorem as saying that if we prescribe in a consistent way how we want an element $\alpha \in K$ to behave under the action of our Galois group – specifically, what element of K it should be multiplied by on applying each element of that group – then we can in fact find an α having this behavior. Now if the Galois group is a cyclic group $\langle\sigma\rangle$, then to describe the behavior of α under this group, it suffices to say what element β it is multiplied by when the generator σ is applied, since its images under elements σ^i can be determined recursively from this datum. We have seen that the consistency condition is then $N(\beta) = 1$. For a general Galois group G, one might similarly specify the behavior of α under G by indicating how it is to be affected by the elements of some generating set for G; but it would be complicated to describe the general form the consistency conditions for these data would take. It is formally simpler just to think of the desired data as specifying for *every* $\sigma \in G$ the element β_σ by which α is multiplied on applying σ. The consistency condition is then easily shown to be

(c39) $$\beta_{\sigma\tau} = \beta_\sigma \, \sigma(\beta_\tau) \qquad (\sigma, \tau \in G).$$

In this section, Lang will prove a generalization of Hilbert's Theorem 90 saying that for every system of β's satisfying (c39), there exists an $\alpha \neq 0$ having the prescribed behavior under G. He will use homological language: a system of elements $(\beta_\sigma)_{\sigma \in G}$ $(\beta_\sigma \in K^*)$ is called a 1-*cocycle* in K^* if it satisfies (c39), and a 1-*coboundary* if it has the form $(\sigma(\alpha)/\alpha)_{\sigma \in G}$ for some $\alpha \in K^*$. The 1-coboundaries form a subgroup of the 1-cocycles, and the first assertion of Theorem VI.10.1 says that the factor-group of the 1-cocycles by the 1-coboundaries is trivial. This is the desired existence result. The second assertion gives the analogous generalization of the *additive* version of Hilbert's Theorem 90.

Some notational points to be aware of: Lang defines the concepts of 1-cocycle and 1-coboundary in a general abelian group on which G acts, which he writes additively, but in the first application, the group will be K^*, which one writes multiplicatively. I have used α and β above in such a way as to match the use of these symbols in the section on Hilbert's Theorem 90; but in this section Lang writes α where I have β.

P.303, second line from bottom [=]: ''so f is a coboundary'' should be ''so $(\tau - 1)f$ is a coboundary''.

Re §VI.11. Non-abelian Kummer extensions. We generally don't have time for these sections in Math 250A; but here are a few errata pointed out by Harry Gindi:

P.305, last line [=]: $\mathbf{Z}(p^{n(p)})$ should be $\mathbf{Z}/p^{n(p)}\mathbf{Z}$.

P.307, 4th-from-last display [=]: The final cy should be ct. On the next line, $\frac{1}{N}\Gamma$ should be Γ'.

P.308, line 5 [=]: $c_M(\Gamma)$ should be $e_M(\Gamma)$.

Re §VI.14. Infinite Galois extensions.

P.313, start of §VI.14 [<]: The first part of this section characterizes the Galois group of an infinite Galois extension K/k in terms of the Galois groups of its finite subextensions F/k. However, the description uses inverse limits (§I.10), which we generally don't look at in Math 250A. The latter part of the section consists mainly of references to advanced work in the literature. One of the basic results of the subject, the characterization of those subgroups of $G(K/k)$ that correspond to intermediate fields, is not

mentioned, because the usual formulation is in topological terms.

Below I give, not a commentary on the section, but ''What I would have put in such section'', namely, a few motivating examples followed by a sketch of these two main results and their proofs, in a form that does not require concepts of topology or of inverse limits. I then point out for the student familiar with those tools how the results proved can be naturally reformulated in terms of them.

(I have a more extensive write-up of the subject, which I generally hand out in Math 250B, which assumes familiarity with the above two concepts, and also with the ring of p-adic integers, and which then develops a characterization of totally disconnected compact topological groups. It can be found at math.berkeley.edu/~gbergman/grad.hndts/infGal+profin.ps .)

Examples of infinite Galois extensions, and two theorems.

To get a feel for the subject, let us begin by determining the automorphism groups of three such extensions K/k. I will first list the examples, then sketch their properties, leaving most of the verification to you. (a) Let $k = \mathbf{Q}$, and adjoin the square roots of all positive rational numbers, equivalently, of all primes, getting $K = \mathbf{Q}(2^{1/2}, 3^{1/2}, 5^{1/2}, ...)$. (b) Let F be a field (of characteristic not 2) containing primitive 2^nth roots of unity for all n, let k be a simple transcendental extension field $k = F(t)$, and look at the Galois extension gotten by adjoining a 2^nth root of t for each n: $K = F(t, t^{1/2}, t^{1/4}, ...)$. (c) Fix a prime p, let $k = \mathbf{Q}$, and let $K = \mathbf{Q}(\zeta_p, \zeta_{p^2}, ..., \zeta_{p^m}, ...)$, where ζ_n denotes a primitive nth root of unity.

In case (a), we see that any $\sigma \in G(K/k)$ is determined by what it does to the elements $p^{1/2}$: some of these elements will be fixed, and others sent to their negatives. With the help of the results in §VI.8 (Kummer Theory), and my comments on that section, you can verify that for any set S of primes, there exists a σ such that $\{p \mid \sigma p^{1/2} = -p^{1/2}\}$ is precisely S. It follows that the Galois group of our extension is isomorphic to the group of all ± 1-valued functions on the set of primes, under pointwise multiplication; that is, to a direct product of countably many copies of Z_2.

In case (b), an element of the automorphism group $G(K/k)$ is similarly determined by what it does to $t^{1/2}$, $t^{1/4}$, etc.; but here the actions are not independent: It can send $t^{1/2}$ to $t^{1/2}$ or $-t^{1/2}$; if it does the former, it must send $t^{1/4}$ to $t^{1/4}$ or $-t^{1/4}$, while if it sends $t^{1/2}$ to $-t^{1/2}$, it must send $t^{1/4}$ to $it^{1/4}$ or $-it^{1/4}$. Likewise, for each behavior of an automorphism on $t^{1/4}$, there are two possible behaviors on $t^{1/8}$, etc.. Associating to each σ the sequence of roots of unity by which the successive elements $t^{2^{-n}}$ are modified, what we have is, first an element of $\boldsymbol{\mu}_2$, then an inverse image of this element in $\boldsymbol{\mu}_4$ under the squaring map $\boldsymbol{\mu}_4 \to \boldsymbol{\mu}_2$, then an inverse image of that element under the squaring map $\boldsymbol{\mu}_8 \to \boldsymbol{\mu}_4$, etc.. Replacing these groups $\boldsymbol{\mu}_{2^n}$ by the isomorphic groups $\mathbf{Z}/2^n\mathbf{Z}$ (e.g., if $F = \mathbf{C}$, identifying $[1] \in \mathbf{Z}/2^n\mathbf{Z}$ with $e^{2\pi i/2^n} \in \boldsymbol{\mu}_{2^n}$), we can say that any element of $G(K/k)$ is determined by a sequence $(..., c_8, c_4, c_2)$ of elements taken from the successive terms of the diagram of groups and surjective homomorphisms

$$(c40) \qquad ... \to \mathbf{Z}/8\mathbf{Z} \to \mathbf{Z}/4\mathbf{Z} \to \mathbf{Z}/2\mathbf{Z},$$

such that each c_{2^i} is the image of the preceding element, $c_{2^{i+1}}$, under the corresponding arrow; and that the group operation of $G(K/k)$ corresponds to componentwise addition of such sequences. (In the language Lang introduces in §I.10, this group is the *inverse limit* of the system of groups and homomorphisms (c40). This particular inverse limit is the additive group of what is called the *ring of 2-adic integers*.) Note that every nonidentity element of this group of sequences has infinite order; for an element whose component at one stage has order 2^i $(i > 0)$ will have component of order 2^{i+1} at the stage immediately to the left, so there is no finite exponent common to all its components. Nevertheless, *locally*, i.e., on each finitely generated subextension $F(t, t^{1/2}, t^{1/4}, ..., t^{2^{-n}})$, each element of this group acts as an automorphism of finite order. Note also that this Galois group, like the preceding, is uncountable.

In case (c), the interested reader can verify, with the help of Theorem VI.3.1, that we have an analogous characterization in terms of the sequence of abelian groups of units in finite rings:

$$\ldots \ \to\ (\mathbf{Z}/p^3\mathbf{Z})^* \ \to\ (\mathbf{Z}/p^2\mathbf{Z})^* \ \to\ (\mathbf{Z}/p\mathbf{Z})^*.$$

One can also show that if p is odd, these groups are isomorphic to the additive groups

$$\ldots \ \to\ \mathbf{Z}/p^2(p-1)\mathbf{Z} \ \to\ \mathbf{Z}/p(p-1)\mathbf{Z} \ \to\ \mathbf{Z}/(p-1)\mathbf{Z}.$$

Using Proposition I.4.3(v), one can break this chain up as the direct product of a constant sequence of cyclic groups, $\mathbf{Z}/(p-1)\mathbf{Z}$, and the sequence

$$\ldots \ \to\ \mathbf{Z}/p^2\mathbf{Z} \ \to\ \mathbf{Z}/p\mathbf{Z} \to 0,$$

which, like (c40), yields an uncountable torsion-free abelian group. So in this case, our Galois group is the product of a finite group and an uncountable torsion-free group. (A similar result is true when $p = 2$, except that the torsion part does not have order $p-1$.)

I have given commutative examples for simplicity and brevity, but you should not find it hard to cook up noncommutative examples as well.

A description of how the Galois group of an arbitrary infinite Galois field extension is related to those of its finite subextensions, generalizing the above examples, is given in the following result, which is straightforward to prove:

Theorem VI.14.c1. *Let K/k be a* (*not necessarily finite*) *Galois extension of fields. Let X denote the set of all subfields $F \subseteq K$ that are* finite *and Galois over k, or, more generally, any subset of that set having the property that every subfield of K finite over k is contained in some member of X.*

Then any automorphism σ of K over k is determined by its restrictions to the members of X. Given a family $(\sigma_F)_{F\in X} \in \prod_{F\in X} G(F/k)$, the necessary and sufficient condition for (σ_F) to arise in the above manner from an element of $G(K/k)$ is the ''consistency condition'', that for every pair of fields $E \subseteq F$ both in X, the restriction homomorphism $G(F/k) \to G(E/k)$ carry σ_F to σ_E.

Thus, $G(K/k)$ can be identified with the subgroup of $\prod_{F\in X} G(F/k)$ consisting of elements $(\sigma_F)_{F\in X}$ satisfying this consistency condition. (*This subgroup of the product is known as the* inverse limit *of the groups $G(F/k)$ with respect to the system of restriction maps.*)

Idea of Proof. The ''consistency condition'' is clearly necessary, and conversely, if it holds for a family (σ_F), then the σ_F, regarded as partial automorphisms of K, fit together to give a total automorphism. □

We now come to a more meaty question: We know from Theorem VI.1.2 that every field L between k and K ''belongs to'' (i.e., is the fixed field of) some subgroup of $G(K/k)$; namely, $G(K/L)$. But which subgroups of $G(K/k)$ ''belong to'' subextensions? The answer is usually given in topological terms, but it can be stated using, instead, the following definition:

Given a set H of automorphisms of K, let us call an automorphism σ of K *H-approximable* if for every finite Galois subextension F of K, there is an element $\tau\in H$ such that $\tau|_F = \sigma|_F$. (Examples: Let $k = \mathbf{Q}$ and $K = \mathbf{Q}(2^{1/2}, 3^{1/2}, 5^{1/2}, \ldots)$, and let H consist of those automorphisms that change the sign of $p^{1/2}$ for only finitely many primes. You should find it easy to verify that *every* automorphism is H-approximable. The same is true for the set H of those automorphisms that change the sign of $p^{1/2}$ for an *even* finite number of primes. But if we let H be the set of automorphisms that change the change the sign of $p^{1/2}$ for a finite number of primes and such that the number of members of $\{2^{1/2}, 3^{1/2}, 5^{1/2}\}$ whose signs are changed is even, then the H-approximable automorphism are just the automorphisms that have the latter property; equivalently, the automorphisms that fix $30^{1/2}$.)

It is straightforward to check that the operation carrying each set H of automorphisms to the set of all

H-approximable automorphisms is a *closure operator* (as defined above in a comment to p.121 of Lang). We shall call that set the *approximation closure* of H, and call H *approximation-closed* if it is its own approximation closure. It is also easy to verify that when H is a subgroup of $G(K/k)$, its approximate-closure is again a subgroup. Let us now prove

Theorem VI.14.c2. *Let K/k be a Galois extension of fields. If H is any subgroup of $G(K/k)$, and K^H its fixed field, then $G(K/K^H)$ i.e., the subgroup belonging to K^H, is the* approximation closure *of H. Hence, the subgroups of $G(K/k)$ belonging to subextensions of K are precisely the approximation-closed subgroups.*

Proof. The last sentence is easily deduced from the preceding sentence, which we shall prove.

First we note that every automorphism σ in the approximation closure of H fixes every element $\alpha \in K^H$. Indeed, given $\alpha \in K^H$, we can find a finite Galois subextension F of K which contains it, and by definition of approximation-closure, there is an element $\tau \in H$ which agrees with σ on F. Since $\alpha \in K^H$, τ fixes α, hence so does σ.

Conversely, suppose $\sigma \in G(K/k)$ fixes all elements of K^H, and let F be any finite Galois subextension of K. We wish to find an element of H which agrees with σ on F. On restricting elements of H to F, we get a group H_F of automorphisms of F over k whose fixed field is $F \cap K^H$. By assumption, the restriction of σ to F fixes all members of this subfield, hence by Artin's Theorem, it must lie in H_F, i.e., there must exist some $\tau \in H$ such that $\tau|_F = \sigma|_F$. Thus, σ is H-approximable, as required. □

Remarks for students familiar with point-set topology. It is not hard to verify that what I have called ''approximation closure'' is the operation of closure under a *topology* on $G(K/k)$. (A basis N of open neighborhoods of the identity in this topology is given by the set of subgroups $G(K/F)$ belonging to the finite Galois subextensions $F \subseteq K$; a basis of open neighborhoods of a general point σ is given by the translate of this family, σN.) If you have seen the concept of an inverse limit of topological spaces, you can verify that the same topology arises by looking at $G(K/k)$ as the inverse limit of the finite *discrete* groups $G(F/k) \cong G(K/k)/G(K/F)$. A finite discrete set is compact, and an inverse limit of compact spaces is compact, so it is a compact topology. Lang refers to it in §VI.14 as the ''Krull topology''.

One may ask *which* topological groups occur as inverse limits of systems of finite groups. These turn out to be those which are compact Hausdorff and totally disconnected, a result proved in the handout mentioned earlier.

There is an interesting open question related to such groups. Note that for any positive integer n, the inverse limit of a system of finite groups all of which have exponent n will again have exponent n. (So, for instance, the Galois group of $\mathbf{Q}(2^{1/2}, 3^{1/2}, 5^{1/2}, ...)$ over $\mathbf{Q}$ has exponent 2.) In particular, an inverse limit of finite groups all having a common exponent will be a group whose elements all have finite order. The question is whether the converse holds. I will pose this question below in two equivalent forms; note that form (b) does not require familiarity either with topological groups or with inverse limits.

Open Question VI.14.c3. (a) *If G is a compact topological group all of whose elements have finite order, must G be of finite exponent? Equivalently,*

(b) *Given a sequence of* finite *groups and surjective homomorphisms, $... \to G_3 \to G_2 \to G_1$, such that the groups G_i have no common finite exponent, can one always find a sequence of elements $(..., g_3, g_2, g_1) \in \prod G_i$ such that for each i, the map $G_{i+1} \to G_i$ takes g_{i+1} to g_i, and such that the g_i have unbounded orders?*

The proof of equivalence of (a) and (b) requires nontrivial results that I will not go into here. (Note that (a) does not assume the group is totally disconnected.) I am told that there is in the literature at least one *incorrect* proof of an affirmative answer to the above open question.

Chapter VIII. Transcendental Extensions.

So far, I only have notes on one section of this chapter:

Re §VIII.1. Transcendence bases.

P.355, beginning of §VIII.1 [<]: If k is a field, X_1, X_2, ... are independent indeterminates over k, and $k(X_1, \dots, X_n) \cong k(X_1, \dots, X_m)$ as k-algebras, must $m = n$? The results of this short section will show that the answer is yes. In fact, they will show that if some algebraic extension of $k(X_1, \dots, X_n)$ is isomorphic as a k-algebra to some algebraic extension of $k(X_1, \dots, X_m)$, then $m = n$. Thus, the number of independent indeterminates involved in this way in an extension K of a field k is an invariant of K.

P.355, display [=]: The notation $M_{(v)}(S)$ is as defined on p.101.

P.355, last four lines [>]: We really want to define the transcendence degree to be the common cardinality of all maximal algebraically independent subsets, but we have not yet proved that these are the same; hence Lang defines it to be the greatest of the cardinalities of such subsets. However, this hedge does not really work: we don't know at this point that there *is* such a greatest cardinality. E.g., it is conceivable that K might have maximal algebraically independent subsets of all finite cardinalities, but none of infinite cardinality, or that there might exist such subsets with cardinalities $\aleph_0$, $\aleph_1$, ... but none with cardinality $\aleph_\omega$. So we really ought to wait till we have proved that all such maximal subsets have the same cardinality before defining the transcendence degree.

The alternative to the term ''transcendence degree'' that Lang gives, ''dimension'', I do not recommend, because it invites confusion with dimension as a vector space over k. (The reason ''dimension'' may sometimes be used is that $k(X_1, \dots, X_n)$ can be thought of as the field of rational functions on ''n-dimensional space'' over k; and, more generally, the transcendence degree of what algebraic geometers call ''the function field associated with an algebraic variety'' is the dimension of that variety.)

Ignore the sentence on the last two lines (''Actually ...''); Lang simply doesn't want to bother with proving the infinite case of the next theorem.

P.356, third paragraph [=]: where Lang says ''**transcendence base**'', I recommend ''transcendence **basis**''.

P.356, statement and proof of Theorem VIII.1.1 [<]: If you have an Addison Wesley printing before the 6th, the proof of this theorem there is incorrect, so be sure to read the proof given in the Errata section of this Companion. That incorporates the five comments marked [=] given immediately below, so in that case you can ignore these.

P.356, sentence containing the first display [=]: This is true because by the sentence before the theorem, w_1 is algebraic over $k(x_1, \dots, x_m)$. Thus it satisfies a polynomial with coefficients in that field, and by clearing denominators we can rewrite this as a polynomial in $m+1$ indeterminates satisfied by $w_1, x_1, \dots, x_m$.

P.356, sentence after first display, ''After renumbering ...'' [=]: This is because the preceding equation must actually involve at least one of the x_i, otherwise it would show w_1 algebraic over k, contradicting the assumption of algebraic independence. By renumbering the x's, we may assume that it involves x_1.

P.356, two lines later [=]: ''Otherwise w_1 would be a root of two irreducible polynomials ...''. Namely, f_1 and the irreducible factor of g_N in question. Note that if this factor vanished on the indicated n-tuple, it would indeed have to involve w_1, so as not to contradict algebraic independence of the x's; also, it cannot be the same irreducible polynomial as f_1, since f_1 involves x_1 and it does not.

P.356, three lines before last display [=]: change ''have found $w_1, \dots, w_r$ $(r < n)$ such that'' to ''have found for some $r < n$ that''.

P.356, last line [=]: The final sentence holds because if we had $n > m$, then the preceding sentence would imply that w_{m+1} was algebraic over $k(w_1, \dots, w_m)$, contradicting algebraic independence of $w_1, \dots, w_n$.

Notes on the above proof. The idea of the above proof is to take a polynomial relation $f_1 = 0$ expressing the algebraicity of w_1 over $k(x_1, \dots, x_m)$, choose any x_i that this relation involves, and then use the same relation to show that x_i is algebraic over the extension generated by w_1 and the other x's, and in this way ''trade off'' x's for w's one by one. However, there is a dangerous trap: If one merely takes any nonzero polynomial f_1 satisfied by w_1 over $k[x_1, \dots, x_m]$, and any x_i that it ''involves'' (i.e., that occurs as a factor of some monomial in f_1), it will not necessarily be true that $f_1 = 0$ constitutes a nontrivial relation satisfied by x_i over the subfield in question. For instance, if $w_1 = x_2 + x_3$, and we take $f_1 = (x_1 - 1)(w_1 - x_2 - x_3)$, then we see that $f_1 = 0$ holds, and f_1 involves x_1, yet as a polynomial in x_1 over $k(w_1, x_2, x_3)$, it is trivial, because of the relation $w_1 = x_2 + x_3$ holding in the latter field. (And in fact, in this situation x_1 is *not* algebraic over $k(w_1, x_2, x_3)$.) So one needs some condition to keep this from happening. The condition Lang uses is that f_1 be irreducible; one that I think gives the nicest proof is to let f_1 be chosen so that the set of x's that it involves is minimal. Another way to get around the problem is to start the proof differently: Where the present proof ''threw in a w'' and showed that we could then ''eject an x'', the proof in Hungerford's text ''pulls an x out'' and shows that one of the w's can ''take its place''. That argument doesn't seem to contain a trap like the above.

More than one text, including printings of Lang before the 6th Addison Wesley printing, give proofs of this theorem that fall into the above trap. I taught from Lang for years and believed the argument; it was only in preparing to teach the same material in an undergraduate course, from another text that used the same argument, and thinking about exactly how to justify it step-by-step for the class, that I saw there was a problem!

P.357, lines 3-4 [>]: Lang says, ''The cardinality statement in the infinite case will be left as an exercise''. That proof uses the same idea as the proof of unique cardinality for bases of non-finitely-generated free abelian groups, given in Lemma I.7.c2 (this *Companion*, comments on Theorem I.7.3). Though there we were concerned with the construction ''the subgroup generated by'', and we here are dealing with the relation ''algebraic over'', the method still works, because if an element w is algebraic over an extension $k(X)$, then there is some finite subset of $X_0 \subseteq X$ such that w is algebraic over $k(X_0)$. You should be able to supply the remaining details.

Now that we know that the transcendence degree of K over k is well defined, we should have a symbol for it. The one I am used to, and will use in class, is $\mathrm{tr.deg.}_k K$.

The other point Lang leaves as an exercise is that one can build up a transcendence basis containing S and contained in Γ. To do this, take a maximal algebraically independent subset $\mathfrak{B}$ of Γ containing S, and note that all elements of Γ, hence all elements of $k(\Gamma)$, hence all elements of K, must be algebraic over $k(\mathfrak{B})$. So $\mathfrak{B}$ is the desired transcendence basis.

P.357, end of §VIII.1 [>]:

Some notes and examples.

If K/k is any field extension, and we take a transcendence basis S for K over k, so that

$$k \subseteq k(S) \subseteq K,$$

then we see that the first of these inclusions is a field-of-rational-functions construction, also called a *pure transcendental extension*, while the second is an *algebraic* extension. Thus, every field extension can be written as an algebraic extension of a pure transcendental extension. However, though the pure transcendental extension $k(S)/k$ is unique *up to isomorphism* (since we have shown that the cardinality of the algebraic independent generating set is unique), it is far from unique. E.g., if $S = \{x\}$, a singleton, then $\{x^2\}$ is another transcendence basis. For another example, if S is any finite set $\{t_1, \dots, t_n\}$, then the set $\{s_1, \dots, s_n\}$ of elementary symmetric polynomials in these generators is also a transcendence basis. (Proof?)

If k is countable, it is easy to show that an extension gotten by adjoining to k countably many

independent indeterminates will again be countable, and we also noted when discussing the construction of algebraic closure that an algebraic extension of a countable field is countable. Hence an extension of a countable field having countable transcendence degree is countable. It follows that the fields of real and complex numbers have *uncountable* transcendence degree over the rationals.

You should find it easy to verify that the field of ''trigonometric functions'' on the real line, i.e., the field of fractions of $\mathbf{R}[\sin x, \cos x]$, has transcendence degree 1 over $\mathbf{R}$. (In fact, it is pure transcendental, i.e., it has the form $\mathbf{R}(f(x))$ for some trigonometric function $f(x)$; but we cannot take this function to be either $\sin x$ or $\cos x$. Can you see why not, and can you find such an f?)

It is not hard to prove that a family of real or complex functions $e^{a_1 x}, e^{a_2 x}, \dots, e^{a_n x}$ is algebraically independent over $\mathbf{R}$, respectively, $\mathbf{C}$, if and only if $a_1, \dots, a_n$ are *linearly* independent over $\mathbf{Q}$. The key step is to verify first that a family of functions $e^{c_1 x}, \dots, e^{c_m x}$ is *linearly* independent over $\mathbf{R}$ or $\mathbf{C}$ if and only if $c_1, \dots, c_m$ are distinct.

You may have noticed a parallelism between the proofs of existence and unique cardinality for bases of vector spaces, and for transcendence bases of field extensions. There is, in fact, a general concept of which both are instances:

General ''dependence structures'', or ''matroids'' (sketch, optional).

Recall that if V is a vector space over a field, then the operator taking each subset $X \subseteq V$ to the subspace that it spans is a *closure operator*, i.e., that if we write the subspace spanned by X as $\mathrm{cl}(X)$, then the operator cl satisfies (c21)-(c23) (this Companion, comment to p.121 of Lang). You should likewise verify that if K is an extension of a field k, then the operator taking a subset $X \subseteq K$ to the field of all elements of K *algebraic* over $k(X)$ is a closure operator. Like most closure operators arising in algebra, these two satisfy the strengthening of (c22) mentioned (but not named) in our earlier discussion,

(c41) $$\mathrm{cl}(X) \;=\; \bigcup_{\text{finite } X_0 \subseteq X} \mathrm{cl}(X_0).$$

A closure operator satisfying (c41) is generally called *finitary*, or *algebraic*,

The above two examples also satisfy a much rarer condition, called the ''exchange property''. This says that for all $X \subseteq S$ and $s, t \in S$,

(c42) $$s \in \mathrm{cl}(X \cup \{t\}) - \mathrm{cl}(X) \;\Rightarrow\; t \in \mathrm{cl}(X \cup \{s\}).$$

(A more colorful term might be the ''You scratch my back, I'll scratch yours'' condition.) You should have no trouble writing down the proof of (c42) in the vector-space case; while in the field-extension case, you should not find it hard to prove (c42) by a modification of the inductive step of the argument I gave to fix Lang's incorrect proof of Theorem I.7.3.

If cl is *any* closure operator on a set S that satisfies (c41) and (c42), then let us say that a subset $X \subseteq S$ *spans* S if $\mathrm{cl}(X) = S$, and that X is *independent* if no $x \in X$ lies in $\mathrm{cl}(X - \{x\})$. Zorn's Lemma, together with (c41), shows that S has a maximal independent subset X. Assuming (c42), one can prove that any such subset spans S; so, calling an independent subset which spans S a *basis*, one has a proof of the existence of bases which includes both the vector-space and the field-extension cases. Moreover, one can prove in this context that any two bases of S have the same cardinality.

A set S with a closure operator cl satisfying (c41) and (c42) is often called a ''dependence system'' by algebraists. The same type of structure was discovered independently by people thinking about the relation of linear dependence among rows of a matrix, and has been extensively studied under the name *matroid*. A good work on the subject is *Matroid Theory* by D. J. A. Welsh, Academic Press (1976). I will use this term below (despite the fact that matroid theorists often restrict their attention to finite matroids, a restriction that I will not make).

Note that if S is a matroid and T a subset of S, then T becomes a matroid under the closure operator $\mathrm{cl}_T(X) = \mathrm{cl}(X) \cap T$; we may call (T, cl_T) a *submatroid* of (S, cl). One of the games in

matroid theory is to see whether various sorts of matroids are isomorphic to submatroids of vector spaces.

For example, given a graph Γ (in the sense of graph theory), there is a classical matroid structure on its edges, under which the closure of a set X of edges consists of those edges e such that the two vertices of e are connected by a path in X. If one maps the set of edges of Γ into a vector space V having the vertices of Γ as a basis, by sending each edge to the difference of its vertices (in either order), then the above matroid structure is induced by that of the vector space V.

For another example, given a vector space V, if we let P be the *projective space* consisting of all 1-dimensional subspaces of V, then P has a natural matroid structure under which the closure of a set of points is the projective subspace that they span. This matroid can be embedded in the matroid structure of V by mapping each point of P to an arbitrary nonzero representative in V. There is also an axiomatic definition of a projective geometry P, which is equivalent to saying that P is a matroid with certain additional properties. It turns out that if $\dim(P)$ (which for geometric reasons is defined in this case to be one less than the number of elements in a basis of P) is > 2, then P will be isomorphic to the projective space associated with a vector space over some division ring. On the other hand, there are known to exist projective planes (2-dimensional projective geometries) which cannot be so represented.

Even the matroid structure associated with algebraic dependence can be embedded in that of a vector space when the characteristic is 0, via the theory of *derivations* (Lang, §VIII.5). However, in positive characteristic it cannot, so far as I know.

Note that the definition of algebraic dependence is based on that of linear dependence: An element α is algebraic over a field k if and only its powers $1, \alpha, \alpha^2, \ldots$ are linearly dependent over k. Now differential equations can be studied algebraically, using the concept of a *differential field* k, that is, a field given with a derivation $x \mapsto x'$. Here the introduction of solutions to differential equations becomes the construction of appropriate differential extension fields. An element f of a differential extension field K/k satisfies some differential equation over k if and only if its derivatives $f, f', f'', \ldots$ are *algebraically dependent* over k. This leads to a theory of *differential dependence* and *differential closure*, which is built upon algebraic dependence in the same way that the latter is built upon linear dependence, and which can again be shown to be a closure operator satisfying (c41) and (c42).

The above are all the examples I know of matroids. The concept, though elegant where it is applicable, only seems to apply to a limited class of natural cases.

(I make use of matroids and their relation with vector spaces in a paper which I consider one of my most interesting, *Constructing division rings as module-theoretic direct limits*, Trans. A. M. S. **354** (2002) 2079-2114.)

Chapter X. Noetherian Rings and Modules.

For this chapter, like the preceding, I have notes on only one section.

Re §X.4. Nakayama's Lemma.

Observe that despite the title of the chapter, rings in this section are *not* assumed Noetherian (first sentence of the section).

P.425, line after first display (or in the Addison Wesley printings, top line of page) [=]: For ''namely a_S'' read ''namely $-a_S$''.

P.425, end of proof of Lemma X.4.1 [>]: More generally, this proof works if we assume E is a left module over a not necessarily commutative ring A, and $\mathbf{a}$ is any 2-sided ideal contained in all maximal *left* ideals of A. For then, at the top of this page, we can say that the *left* ideal generated by $1+a$ is not proper, hence $1+a$ is left invertible, allowing us to complete the proof.

Remark: The intersection of all maximal left ideals of a ring coincides with the intersection of all maximal right ideals, and is thus a 2-sided ideal, called the (*Jacobson*) *radical* of A. So the assumption

on $\mathbf{a}$ is really that it is contained in this radical. (Some properties of the Jacobson radical are developed in Exercises XVII:L1-5, p.661.)

Here is an alternative proof of the lemma, which I like. Obviously, it suffices to prove that a *nonzero* finitely generated module E satisfies $\mathbf{a}E \neq E$. Now because E is finitely generated, it has a maximal proper submodule F. (Do you recall how this is proved?) Thus, E/F is a simple module; hence the annihilator of every element of this module is a maximal left ideal of A. Therefore, $\mathbf{a}(E/F) = 0$, so $\mathbf{a}E \subseteq F$, so $\mathbf{a}E \neq E$.

The lemma is false without the finite generation hypothesis. For instance, if we take for A any local integral domain that is not a field, for $\mathbf{a}$ its maximal ideal, and for E its field of fractions, then $\mathbf{a}E = E$, but $E \neq 0$.

The next two lemmas in Lang, which both say virtually the same thing, modify the statement of the preceding lemma simultaneously in two ways: First, they apply it to a factor-module E/F, to get a statement that under certain conditions, a submodule $F \subseteq E$ must be all of E; and second, they assume A local, so that without loss of generality $\mathbf{a}$ can be taken to be the maximal ideal. These two modifications can, of course, each be made separately.

These and the remaining results of this section are also true for noncommutative A. (For the definition of a noncommutative local ring, see Exercise II.4:3.) Unfortunately, in noncommutative ring theory one cannot in general get local rings from arbitrary rings by localization, so these results, though still important, are not as powerful a tool in studying general noncommutative rings as in the commutative case.

P.425, proof of Theorem X.4.4 [=]: Where Lang says ''The first statement'', he means the two sentences, ''*Then E is free. In fact, ...*''. Likewise, by ''the second statement'' (bottom line of this page) he means the last sentence of the theorem.

Chapter XII. Absolute Values.

We don't cover this in Math 250, but Johann Schuster has sent me one erratum:

Re §XII.1. Definitions, Dependence, and Independence.

P.466, third display [=]: The inequality should be ''<''.

Also, the final conclusion of the proof would be clearer if, after defining λ earlier, one had noted that this is the real number such that $|x_0|_2 = |x_0|_1^{\lambda}$. At the end, one combines this with the definition of α as the real number such that $|x|_1 = |x_0|_1^{\alpha}$, and the equality one has just proved.

Chapter XIII. Matrices and Linear Maps.

[These notes on Chapters XIII, XIV, and XVI were written after teaching from this material in Spring 1997. Until I get to teach from it again and polish them further, they are liable to be somewhat rough. As with everything else in this Companion, I welcome corrections to these notes, suggestions on points that could use clarifying, etc..]

Re §XIII.1. Matrices.

P.504, first display [>].

Some background and some notation.

The definitions Lang has just given, of matrices and their additive and multiplicative structures, are of course standard. But where do they come from?

Consider a linear map f from the standard free module of rank n, R^n, to the standard free module of rank m, R^m. If we write the natural bases of these modules $\{e_1, \ldots, e_n\}$ and $\{e_1, \ldots, e_m\}$, then

$f\colon R^n \to R^m$ is determined by the n elements $f(e_1), \dots, f(e_n)$, each of which is in turn determined by the coefficients with which the m elements $e_1, \dots, e_m$ occur in it. Thus, writing $f(e_j) = \Sigma\, a_{ij}\, e_i$, we see that f is determined by the mn elements $a_{ij} \in R$ $(1 \leq i \leq m,\ 1 \leq i \leq n)$. It is natural to arrange these in a rectangular array, so that those coefficients expressing the image of a given e_j form a column, while those occurring as coefficients of a given e_i in these images form a row. The laws of addition and multiplication of matrices express addition and composition of linear maps in terms of these coefficients.

There are other standard uses of matrices: to describe changes of bases of free modules, to represent bilinear forms, etc.. When these come up below, however, we shall see how to relate them all to the above interpretation.

Note that the hom-set $\mathrm{Hom}(R^1, R^m)$ can be identified with R^m by associating to each homomorphism the element to which it sends the basis element $e_1 = 1 \in R$. This has the pleasant consequence that rather than setting up distinct notations for matrices and for elements of the free modules on which they act, we can treat these elements as special cases of matrices, namely as matrices with only one column.

There is a convenient notation, though not used by Lang, in which one writes the set of $m \times n$ matrices over a ring R as ${}^mR^n$. One also generally drops m and/or n if its value is 1. I will use this notation for the remainder of the notes on this chapter. In particular, the modules denoted by R^n and R^m above will henceforth be written as ${}^nR$ and ${}^mR$.

Note that ${}^nR^n$ is the ring Lang writes $\mathrm{Mat}_n(R)$. Matrix multiplication makes the set ${}^mR^n$ a left module over ${}^mR^m$, and a right module over ${}^nR^n$. This multiplication corresponds to composing linear maps ${}^nR \to {}^mR$ on the left with endomorphisms of ${}^mR$ and on the right with endomorphisms of ${}^nR$. Associativity of composition of functions implies that the *left* module structure acts by endomorphisms of ${}^mR^n$ as a *right* module, and vice versa. An abelian group with such a pair of module structures is called an $({}^mR^m, {}^nR^n)$-*bimodule*. I will occasionally mention bimodules parenthetically below, but will not introduce the concept formally until a later chapter (discussion entitled *Tensor products in the noncommutative context*, at end of comments to §XVI.1).

So far, I have tacitly followed Lang and assumed R commutative. If we drop this assumption (though always assuming our rings have unity), what we have said above works, as long as we regard ${}^mR$ as a free *right* R-module. (So where I wrote $f(e_j) = \Sigma\, a_{ij}\, e_i$, this should read $\Sigma\, e_i\, a_{ij}$.) Indeed, its right R-module structure is the $n=1$ case of the right ${}^nR^n$-module structure on ${}^mR^n$.

If M and N are right modules over a noncommutative ring R, then $\mathrm{Hom}(M, N)$, though it is an abelian group, does not form an R-module, since when one multiplies a homomorphism f by an element $a \in R$, the resulting map $x \mapsto f(x)a$ does not in general respect multiplication by elements of R: $f(xr)a = f(x)ra \neq f(x)ar$. However, when $N = R$, it makes sense to *left*-multiply elements $f(x)$ by members of R, and it is easy to verify that this makes $\mathrm{Hom}(M, R)$ a left R-module. (More generally, if M is a right S-module and N an (R, S)-bimodule, then the abelian group of right S-module homomorphisms $\mathrm{Hom}(M, N)$ acquires in this way a natural structure of left R-module.) It is easy to verify that if M is a free right module of finite rank with basis B, then $\mathrm{Hom}(M, R)$ is free as a left module on the ''dual basis''. If we take for M the standard free right module ${}^nR$, we find that $\mathrm{Hom}(M, R) = \mathrm{Hom}({}^nR, {}^1R) = {}^1R^n = R^n$. That is, the set of *row* vectors of length n has a natural structure of free *left* module of rank n, and we may take it as our ''standard'' free left module of that rank. Its left module structure is given by the action of 1×1 matrices on the left of row vectors.

Since the rules for matrix multiplication are symmetric, ${}^mR^n$ can be regarded equally as representing the right module homomorphisms ${}^nR \to {}^mR$ and the left module homomorphisms $R^m \to R^n$. Note, however, that in using these conventions, we must understand homomorphisms of right modules to be written to the *left* of their arguments, while homomorphisms of left modules are written to the *right* of their arguments, with each sort being composed in the way natural to the side on which it is written: $A(Bx) = (AB)x$ while $(yA)B = y(AB)$. (This is a convention that is often followed in noncommutative ring theory generally: Homomorphisms among right modules are written to the left of their arguments, and vice versa.

Note that it makes the equations saying that such homomorphisms respect scalar multiplication into formal associative laws: $f(xr) = (fx)r$, respectively $(rx)f = r(xf)$ for f a homomorphism, x a module element, and r a ring element.)

Of course, if R is commutative, it does not mean much to say ''the set ${}^nR$ of column vectors represents the free right module on n generators while the set R^n of row vectors represents the free left module''. In that case, we simply regard ${}^nR$ and R^n as two copies of the free R-module of rank n, subject to different notational conventions regarding linear maps.

Now back to Lang.

P.505, paragraph near top, defining the **transpose** of a matrix [~]: Although, as I have noted, the definitions of matrix addition and multiplication are as useful in studying modules over noncommutative rings as in the commutative case, the concept of the transpose of a matrix is only natural when R is commutative. In that case, it can be motivated by taking the abovementioned isomorphism of (right = left) modules ${}^nR \cong R^n$, and noting that this leads to an identification between $\mathrm{Hom}({}^nR, {}^mR) \cong {}^mR^n$ and $\mathrm{Hom}(R^n, R^m) \cong {}^nR^m$, which we denote ^{t}. The difference between the notations for composition of these two sorts of morphisms leads to the identity ${}^t(AB) = {}^tB\,{}^tA$. For R noncommutative, both the motivation and the identity fail; indeed, the 1×1 case of the identity is the condition of commutativity.

Nevertheless, the symbol ^{t} is sometimes used in noncommutative ring theory for typographical convenience: If one wants to refer to the column with entries $x_1, \dots, x_m$, without leaving a lot of space between lines, one can write it ${}^t(x_1, \dots, x_m)$.

Two other definitions that are natural only for matrices over commutative rings are those of the *trace*, defined later on this page of Lang, and the *determinant*, introduced in §XIII.4.

In general, when definitions or results in Lang apply only for R commutative, I will say so; likewise, I will point when there are nontrivial adjustments to be made in extending results to the noncommutative context. When I refer to a result in Lang as translating to, say, a certain statement about right modules, it should be understood that the obvious corresponding statement about left modules is also true, and may indeed be called on in subsequent arguments. Where I say nothing to the contrary, assume the results and proofs he gives are valid in the noncommutative case, perhaps with some straightforward notational adjustment. (But if you find some difficulty, let me know – I may have missed a point.)

P.505, middle [=]: For ''A similar remark'' read ''This definition of φA''.

P.505, bottom [>]. *Further remarks.* There is a kind of matrix notation more general than those we have discussed here. (We mentioned it briefly in our discussion of §III.5.) Rather than considering the free right R-modules ${}^nR$ and ${}^mR$, suppose we look at any two right R-modules given with finite direct-sum decompositions: $N = N_1 \oplus \dots \oplus N_n$, $M = M_1 \oplus \dots \oplus M_m$. A homomorphism $f: N \to M$ is clearly determined by its restrictions $f_j: N_j \to M$ to the n summands of N, and each of these is determined by its composites with the projections of M onto its m components $f_{ij}: N_j \to M_i$. If we write these mn homomorphisms in a rectangular array, it is easy to see that the formal laws of matrix addition and multiplication correctly describe addition and composition of maps among such direct sums of modules. In particular, the ring of endomorphisms of a module of the form $N = N_1 \oplus \dots \oplus N_n$ may be expressed as a square ''matrix ring'', with homomorphisms among these modules as the matrix entries.

There are still other sorts of generalized matrices that are occasionally useful, which I won't go into here. For instance, if B is an (R,S)-bimodule, one finds that formal matrix operations make the set $\left(\begin{smallmatrix} R & B \\ 0 & S \end{smallmatrix}\right)$ a ring.

Re §XIII.2. The rank of a matrix.

P.506, first paragraph [>]: The result to be proved here, that the row and column rank of a matrix over a field are equal, is true, more generally, for matrices over a division ring, where, of course, one defines the row rank of such a matrix as the dimension of the *left* vector space spanned by the rows, and the column rank as the dimension of the *right* vector space spanned by the columns. The same proof works, if we assume the duality theory of §III.6 verified for modules over not necessarily commutative rings, the dual of

a right module being regarded as a left module and vice versa.

However, let me give a proof that does not call on duality theory. We will need

Definition XIII.2.c1. *Let R be a ring, m and n nonnegative integers, and $A \in {}^mR^n$. Then the* inner rank *of A is defined to be the least nonnegative integer r such that A can be factored $A = BC$, with $B \in {}^mR^r$ and $C \in {}^rR^n$.*

(*Here we allow the possibility $r = 0$, understanding an $m \times 0$ or a $0 \times n$ matrix to have no entries, and to represent the unique homomorphism $\{0\} \to {}^mR$, respectively ${}^nR \to \{0\}$. Clearly, A has rank 0 if and only if it is the zero $m \times n$ matrix.*)

Now a matrix A has column rank r if and only if all of its columns can be expressed as right linear combinations of some r column vectors, $b_1, \dots, b_r$, but cannot be so expressed using fewer. But such an expression is equivalent to a factorization $A = BC$ where B is the matrix with those r vectors as its columns. Thus, the column rank and the inner rank of A are equal. Looking at rows instead of columns, we similarly deduce that the row rank and the inner rank are equal. Thus we have

Lemma XIII.2.c2. *The row rank, the inner rank, and the column rank of a matrix over a division ring are all equal.* □

Incidentally, the definitions of row and column rank are based implicitly on the fact that a vector space has a well-defined dimension, i.e., that all its bases have the same number of elements. This was proved for vector spaces over fields in §III.5. As I noted there, the proof is equally valid for vector spaces over division rings.

Re §XIII.3. Matrices and linear maps.

P.507, first paragraph of §XIII.3 [>]: This shows that a basis $\xi_1, \dots, \xi_n$ of a right module E determines an isomorphism ${}^nR \to E$. This isomorphism can be characterized as the unique module homomorphism sending the ith standard basis vector e_i of ${}^nR$ to ξ_i. Inversely, any isomorphism ${}^nR \to E$ determines a basis of E indexed by $\{1, \dots, n\}$, namely the image under the homomorphism of the standard basis of ${}^nR$. Thus, a choice of an n-element ordered basis for E is *equivalent* to an isomorphism ${}^nR \to E$. This observation will be a key tool below in translating results about arbitrary free modules into matrix language.

P.507, second paragraph of §XIII.3 [>]: This proof that a free module over a commutative ring has unique dimension uses the fact that vector spaces (i.e., modules over *fields*) have unique dimension, and that every commutative ring admits a homomorphism into a field. Cf. this Companion, discussion at end of §III.5, which also shows how to construct noncommutative rings with free modules having *nonunique* ranks. (Apologies for having written homomorphisms of left modules on the left of their arguments there; but I didn't want to introduce unfamiliar notation for that brief digression.)

Despite Lang's use of the word *dimension*, it is more usual to speak of the *rank* of a free module, if the base ring is not a division ring, and I will do so below.

P.508, Proposition XIII.3.1 [~]: Note that the column vectors considered here are n-tuples of elements of a free module, which is itself of rank n. Moreover, since a matrix $A \in {}^nR^n$ acts on the left on elements of ${}^nE$, we should, in formulating this result for R not necessarily commutative, take E to be a left R-module.

In particular, if we take for E the standard free left R-module of rank n, consisting of row vectors, we find that our column vectors of elements of E are height-n columns of length-n rows; in other words, they look like elements of ${}^nR^n$ itself. If we take for $x_1, \dots, x_n$ the standard basis of R^n, then the column that they form is the identity matrix, and the left-hand side of the displayed equation is just A. So this case of the proposition says that A is invertible if and only if its rows form a basis of R^n.

Some quick exercises in the use of the proposition: Suppose V is a 3-dimensional k-vector-space with

basis $\{x, y, z\}$. Under what conditions, if any, on k will $\{x+y, y+z, z+x\}$ be a basis of V? What about $\{x-y, y-z, z-x\}$? If $\mathfrak{B}$ is a $\mathbf{Z}/p\mathbf{Z}$-vector-space basis of $^n(\mathbf{Z}/p\mathbf{Z})$, and the matrix having the elements of $\mathfrak{B}$ for rows has determinant other than $\pm 1 \in \mathbf{Z}/p\mathbf{Z}$, show (assuming standard properties of the determinant, though we will not formally see these till the next section) that $\mathfrak{B}$ is not the image of any basis of the free abelian group $^n\mathbf{Z}$ under the natural map $^n\mathbf{Z} \to {}^n(\mathbf{Z}/p\mathbf{Z})$.

P.509, bottom [>]: As noted earlier, a choice of an n-element basis for a right R-module E induces an isomorphism between ^nR and E. Hence given bases $\mathfrak{B}$ and $\mathfrak{B}'$ of right modules E and F, say with n and m elements respectively, a module-homomorphism $f: E \to F$ is equivalent to a homomorphism of standard free modules:

$$\begin{array}{ccc} {}^nR & & {}^mR \\ \downarrow\mathfrak{B} & & \downarrow\mathfrak{B}' \\ E & \xrightarrow{f} & F. \end{array}$$

Here the arrows marked $\mathfrak{B}$, $\mathfrak{B}'$ represent the module *isomorphisms* taking the standard bases of ^nR and ^mR to these bases of E and F. The resulting homomorphism $^nR \to {}^mR$ corresponds to a matrix $A \in {}^mR^n$, which Lang denotes $M_{\mathfrak{B}'}^{\mathfrak{B}}(f)$. If you draw the appropriate diagrams, you will now see that the equation in the proposition at the beginning of the next page is immediate! Likewise, given two bases $\mathfrak{B}$ and $\mathfrak{B}'$ of a right module E (say of n and m elements respectively, where $m = n$ if R is commutative), we get a diagram

$$\begin{array}{ccc} {}^nR & & {}^mR \\ \downarrow\mathfrak{B} & & \downarrow\mathfrak{B}' \\ E & = & E. \end{array}$$

from which we see that $M_{\mathfrak{B}'}^{\mathfrak{B}}(\mathrm{id})$ represents the linear map $^nR \to {}^mR$ that carries the expression for an element of E with respect to the basis $\mathfrak{B}$ to its expression with respect to the basis $\mathfrak{B}'$.

P.510, to end of section [~]: If it is hard to see what the first two corollaries on p.510 are asserting, this is because Lang has included the proofs as part of the statements. The real assertion of Corollary XIII.3.4 is that a change-of-basis matrix $M_{\mathfrak{B}'}^{\mathfrak{B}}(\mathrm{id})$ is invertible; the displayed equation is the proof. Likewise, the point of Corollary XIII.3.5 is that the left side of the display equals the right side. The middle term gives the proof.

Everything in this section is valid for noncommutative as well as for commutative rings, except the last sentence of Corollary XIII.3.5 (since for noncommutative R one cannot speak of R-algebras), and the remark about the trace function at the very end of the section.

Re §XIII.4. Determinants.

P.511, start of the section [<]:

I will assume in my comments on the remaining sections of this chapter that R is commutative.

(There is a concept of multilinear map appropriate to noncommutative ring theory, which I will introduce in the notes on §XVI.1, but there is no natural concept of *alternating* multilinear maps even in that context.)

P.514, line 3 [=]: ''Using a previous lemma ...'' – namely, Lemma XIII.4.3. On the same page, two lines above the corollary, ''by the lemma'' again refers to Lemma XIII.4.3.

P.516, first sentence beginning after middle display [~]: Another proof of this result is noted in the comments to §VI.4 in this Companion.

P.517, line 1 [>]: In general, I favor more mnemonic notation; so where Lang writes $L_a^n(E, F)$, I would prefer something like $\mathrm{Lin}_{\mathrm{alt}}^n(E, F)$. Of course, when in the creative throes of discovery, use in your scratch-work whatever notation you are comfortable with. But in a good write-up, try for notation whose meaning is fairly transparent, if you can do so without making it excessively bulky. In this Companion, I will follow Lang's notation $L_a^n(E, F)$.

P.518, Proposition XIII.4.16 [=]: Note that this is essentially Cramer's rule (Theorem XIII.4.4),

restated.

P.520, Proposition XIII.4.19 [=]: Note that this proposition characterizes $D(f)$ in a way that does not depend on a matrix representation of f, and hence a choice of basis of E.

There is a slightly more abstract, but very elegant variant to this characterization of the determinant, in which one associates to every module E and nonnegative integer d a module $\bigwedge^d E$, the ''dth exterior power of E'', having a universal d-linear alternating map of $E\times\ldots\times E$ into it. One finds that if E is free of rank n, then $\bigwedge^d E$ is free of rank $\binom{n}{d}$. In particular, $\bigwedge^n E$ is free of rank 1. The construction $\bigwedge^d$ is a functor, hence any endomorphism f of E induces an endomorphism $\bigwedge^n f$ of this rank-1 free module, which must be given by multiplication by an element $D(f)\in R$. Some of the theory of the functors $\bigwedge^d$ (and of $\bigwedge = \bigoplus_d \bigwedge^d$, which is not only module- but algebra-valued) is developed by Lang in Chapter XIX (not presently covered in this Companion); cf. also a handout of mine, *Tensor algebras, exterior algebras, and symmetric algebras*, accessible through my web page.

P.521, second line of second paragraph [=]: For ''view it multiplicatively, as'' read ''restrict it to automorphisms, getting''.

P.522, end of §XIII.4 [>]. Here is a curious question: Suppose k is a field and n a positive integer, and we form the polynomial ring in n^2 indeterminates, $k[x_{11},\ldots,x_{nn}]$, and let $X = (x_{ij})$ be the $n\times n$ matrix formed from these indeterminates. Then as shown in Proposition XIII.4.16, we get a factorization (for which I will use different notation from Lang): $\det(X)\,I = X\,\mathrm{adj}(X) = \mathrm{adj}(X)\,X$. Is any further factorization possible, and if so of what sort?

This question is answered completely for k of characteristic 0 in *Transformation Groups*, **11** (2006) pp.7-15; but for prime characteristic it is still open.

Re §XIII.5. Duality.

P.522, last six lines [>]: Where Lang writes **perpendicular**, the more common term is ''orthogonal''. Note that the notation $S^\perp$ is unambiguous if E and F are disjoint; but if, for example, they are the same space, then for a subset $S\subseteq E$, the symbol $S^\perp$ could mean either $\{x\in E \mid (\forall\, s\in S)\ s\perp x\}$ or $\{x\in E \mid (\forall\, s\in S)\ x\perp s\}$; and if f is neither symmetric or antisymmetric, these may be distinct. In setting up the abstract theory, as is done here, one can treat these spaces as though they were disjoint without actually making this a hypothesis, since in any given context a set S will be described as a subset of E or as a subset of F, so that the sense of $S^\perp$ will be clear. But if one applies this theory to a particular case in which E and F are not disjoint, one may need to introduce distinct notations for the above sets. This sort of situation is not uncommon in mathematics.

P.523, first boxed display [=]: If we adopt ''exponential notation'', writing B^A for $\mathrm{Hom}(A,B)$, and using the *ad hoc* notation $C^{A\cdot B}$ for $L^2(A,\,B;\,C)$, then (replacing the R in this display by a general module D, and interchanging the roles of E and F on the right, which is OK because we are still assuming our ring commutative, so that the construction $L^2(E,\,F)$ is symmetric in E and F), the display takes the suggestive form $D^{E\cdot F}\cong (D^E)^F$. (In noncommutative ring theory, one can formulate a generalization valid for bimodules, but we won't look at it here.)

P.523, last paragraph [>]: Actually, this only needs f to be nonsingular on the left. For in that case, we can apply the isomorphism $\mathrm{Hom}_R(F,R)\cong E$ to the previous boxed display to turn the right-hand side into $\mathrm{Hom}_R(E,E) = \mathrm{End}_R(E)$. The top three paragraphs on the next page just make this construction explicit.

P.524, 7 lines from bottom [<]: Suppose we are given two modules E and F, together with right-nonsingular bilinear forms relating them to modules E' and F':

$$E\times E'\to R, \qquad F\times F'\to R.$$

These forms allow us to identify E' and F' with $E^\vee = \mathrm{Hom}_R(E,R)$ and $F^\vee = \mathrm{Hom}_R(F,R)$ respectively. Now any homomorphism $A\colon E\to F$ induces a homomorphism $A^\vee\colon F^\vee\to E^\vee$, hence, via the above identifications, a homomorphism $F'\to E'$, which we can call ${}^tA$. The preceding discussion has concerned the case where $E = F'$ and $F = E'$. The paragraph beginning here concerns the case

where $E = E'$ and $F = F'$. (So it is not, as Lang claims, ''more general'' than the preceding case.)

P.525, **Note** following Proposition XIII.5.1 [>]: This is true because ${}^tAA = \mathrm{id} \Rightarrow A$ has invertible determinant $\Rightarrow A$ is invertible.

One may ask whether the same implication holds without the assumption that E is free of finite rank. Here is an almost-counterexample, where the given bilinear form is non-degenerate, though not, as in the proposition, nonsingular. Let R be a field, E a countable-dimensional vector space with basis $\{x_0, x_1, \ldots\}$, and let our bilinear form be given by $\langle x_i, x_j \rangle = \delta_{ij}$. Let $Ax_i = x_{i+1}$. Knowing only that $\langle\ ,\ \rangle$ is non-degenerate, we do not know a priori that there exists an operator ${}^tA$ related to A by the equation $\langle Ax, y\rangle = \langle x, {}^tAy\rangle$; but in this case there is such an operator, given by ${}^tAx_{i+1} = x_i$ $(i \geq 0)$, ${}^tAx_0 = 0$. We find that ${}^tAA = \mathrm{id}$ but that $A\,{}^tA \neq \mathrm{id}$.

To get an example where $\langle\ ,\ \rangle$ is in fact nonsingular, one must go out of the realm of vector spaces, since for finite-dimensional vector spaces the implication is true, while the dual of an infinite-dimensional vector space E is always of larger dimension than E, hence cannot be isomorphic to E. Here are some hints for the student interested in constructing a module-theoretic counterexample. Note that for any module E, there is a natural homomorphism $E \to E^{\vee\vee}$. If this is an isomorphism, then $E \oplus E^{\vee}$ will be isomorphic to its dual, hence will admit a non-singular bilinear form. Now try your hand at Exercise I.7:1(b); a solution to this, together with the above observations, will provide the tool needed to find the desired example with $R = \mathbf{Z}$.

Re §XIII.6. Matrices and bilinear forms.

In this and the remaining sections of this chapter, we shall adopt Lang's assumption that R is commutative. (After going through the first few sections of Chapter XVI and my comments on these, you might want to revisit this section and think about how to give generalizations of its ideas to free right and left modules over general rings.)

P.527, display at bottom [>]: This shows that a bilinear form f can be described by a matrix g_{ij}. This is the matrix (with respect to the given basis of E, and the basis of $F^{\vee}$ dual to the given basis of F) of the linear map $E \to F^{\vee}$ which takes $x \in E$ to $f(x, -) \in F^{\vee}$. Thus, the use of matrices to describe bilinear forms reduces to that of describing linear maps.

Re §XIII.7. Sesquilinear duality.

P.533, definition of **sesquilinear form** [>]: Clearly, this is just a form that is *bilinear* with respect to the given module structure on E and a modified module structure on F (which Lang will call, a few paragraphs from now, the ''anti-module'' structure). So in this generality, there is nothing new. It is only on adding the condition $E = F$ (in results XIII.7.{1,5,6}) that one gets a theory that doesn't reduce to the corresponding theory for bilinear forms.

P.534, line preceding Proposition XIII.7.2 [<]: Ignore this line; or let me know if you can figure out what it means!

Also, the last line on this page should be indented like the three that precede it, being one of the equivalent conditions.

Re §XIII.8-9. No notes on these at present.

Chapter XIV. Representation of One Endomorphism.

Re §XIV.1. Representations.

P.553, bottom [>]: A ''representation of R in E'', as here defined, is the same as a left R-module structure on E extending the given k-module structure. This alternative language probably comes from the theory of group representations, where actions of a group G by linear automorphisms on a vector space V, which are equivalent to kG-module structures on V, were studied before the concept of module was defined, so that the group-theorists developed their own language. Hopefully, this duplication of terminology will eventually be eliminated.

P.554, second paragraph [=]: In the last two sections of this chapter, the ring R will in fact be $k[x]$. But the definitions in this section should be understood for general R, not necessarily commutative.

Re §XIV.2. Decomposition over one endomorphism.

P.556, near bottom, definition of **principal** [>]: What Lang calls a ''principal module'' is more often called a ''cyclic module'' (cf. the definition of a cyclic group), and I will use the latter term.

Remarks: If E is a cyclic left module over any ring R, with generator v, then the *annihilator* of v in R, $\mathrm{Ann}_R v = \{r \in R \mid rv = 0\}$, is a left ideal of R, and it is easy to see that $E \cong R/\mathrm{Ann}_R v$ as a left R-module. (This is analogous to the description of the general transitive G-set in terms of the isotropy subgroup of an element.) If R is *commutative*, then any element of $\mathrm{Ann}_R v$ also annihilates every element $av \in E$, hence $\mathrm{Ann}_R v$ can also be characterized as $\{r \in R \mid \forall\, x \in E,\ rx = 0\}$, i.e., as the annihilator of the whole module E, equivalently, as the kernel of the given map $R \to \mathrm{End}(E)$, and so is independent of a choice of cyclic generator. This is not true in the noncommutative case, where the annihilator of the whole module will be a 2-sided ideal, while the annihilator of a single element can be any left ideal. (Given $a \in R$, you should find it easy to write down the formula for $\mathrm{Ann}_R av$ in terms of $\mathrm{Ann}_R v$.) But for R commutative, we thus see that the isomorphism classes of cyclic R-modules E are classified by the ideals of R. If R is a principal ideal domain, then these ideals are in turn classified by elements of R, up to unit factors. In the case $R = k[t]$, such ideals, when nonzero, correspond to monic polynomials $q(t)$.

Given a monic polynomial $q(t)$, it is easy to write down a matrix for the endomorphism of the $k[t]$-module $k[t]/q(t)$ given by multiplication by t; this is done on the next page; the result is generally called the *companion matrix* of $q(t)$ (but is not named by Lang). As Lang then notes, the structure theorem for finitely generated modules over principal ideal domains allows one to reduce the study of arbitrary finitely generated modules over such rings to that of cyclic modules, and hence to classify $k[t]$-modules finite-dimensional over k, i.e., finite-dimensional k-vector-spaces given with endomorphisms A, in terms of families of monic polynomials.

P.557, Corollary XIV.2.2 [<]: We now want to translate our classification of vector spaces given with an endomorphism into a classification of matrices. Two pairs $(k^{(n)}, A)$ and $(k^{(n)}, B)$ (using Lang's notation here instead of my ${}^n k$) are isomorphic if and only if some change-of-basis transformation converts A into B; equivalently, if and only if $B = CAC^{-1}$ for some invertible matrix C; so Theorem XIV.2.1, which shows that every pair (E, A) is isomorphic to a unique member of a certain family of such pairs, shows that every square matrix over k is conjugate to a unique member of a certain family of matrices; namely, those given by diagonal blocks of companion matrices of a sequence of monic polynomials in which each polynomial is divisible by the next.

Now if two matrices are conjugate over k to *distinct* members of this family, they are still conjugate to those distinct matrices over any larger field k'; hence by the unicity part of the above result, they cannot be conjugate to one another over this extension field. This is the result of Corollary XIV.2.2. The last sentence of the proof is vaguely worded because Lang has not introduced the term ''companion matrix'', which would allow him to state it more precisely.

P.558, Theorem 2.4 [<]: As Lang has just noted, we have another structure theorem for finitely generated modules over a principal ideal domain, from which one can get a different canonical form,

expressing a matrix as a diagonal sum of companion matrices of powers p^e of monic irreducible polynomials.

But for such polynomials there is in fact a matrix representation nicer than the companion matrix. Observe that the submodules of the $k[t]$-module $k[t]/(p^e)$ form a chain

$$(1)/(p^e) \supseteq (p)/(p^e) \supseteq \ldots \supseteq (p^e)/(p^e) = \{0\}.$$

A k-basis for the first term of this chain modulo the second is given by $1, \ldots, t^{d-1}$, where d is the degree of p; a basis for the second modulo the third is given by $p, tp, \ldots, t^{d-1}p$, and so on. Taking the union of these bases, we get a basis for the whole module $k[t]/(p^e)$, in terms of which the matrix for multiplication by t consists of e diagonally arranged copies of the companion matrix for p, together with single entries ''1'' just below and to the left of the corners where these blocks touch, representing the fact that when one multiplies $t^{d-1}p^i$ by t, and subtracts off an appropriate linear combination of lower terms $t^j p^i$, one gets, not 0, but p^{i+1}. (I suggest you work out the details.)

The form of these matrices is especially simple when p is a linear polynomial $t-\alpha$, and is given in Theorem XIV.2.4. If k is algebraically closed, then every irreducible polynomial is linear, and the resulting canonical form for matrices is the well-known Jordan Canonical Form (Corollary XIV.2.5, p.559).

P.560, Corollary XIV.2.7 [>]: Actually, this can be proved more easily, without the assumption that K is separable (or even commutative), and without the use of the preceding theorem. Given a finite-dimensional division algebra K over k, and two representations of K on V, let us denote by V_1 and V_2 the corresponding K-vector-spaces. The K-dimension of each is clearly $(\dim_k V)/[K:k]$, hence, having the same dimension, they are isomorphic as K-vector-spaces, which is equivalent to the stated conclusion.

Re §XIV.3. The characteristic polynomial.

P.561, second line after second display [=]: R should be k. (And for ''in'' I would say ''over'', both here and three lines before that display.)

P.561, five lines above Theorem XIV.3.1 [=]: Why do we define the characteristic polynomial of the unique endomorphism of the zero-module to be 1? This is a particular case of a more general convention Lang should have made earlier, that the *determinant* of the unique endomorphism of the zero module is 1. A pragmatic justification for this is that it makes various formulas come out right; e.g., given endomorphisms e and f of free modules E and F, one gets an endomorphism $e \oplus f$ of the module $E \oplus F$. When these modules have positive rank, one sees that the determinant of $e \oplus f$ is the product of the determinants of e and of f. To make this true if E or F is the trivial module, one should define the determinant of its unique endomorphism to be 1.

One can also justify this formally, by examining closely what ''n-linear alternating map'' means when $n = 0$. One finds that it is equivalent to a choice of element of the codomain module, and that the unique endomorphism of the zero module has no effect on the module of 0-linear alternating forms, $L_a^0(\{0\}, k) \cong k$, i.e., it acts on it by multiplication by 1. I leave the details to the interested reader.

P.561, third from bottom line [=]: Change ''Let $\tilde{B}(t)$ be'' to ''Let $B(t) = tI_n - M$ and let $\tilde{B}(t)$ be''.

Pp.561-562, proof of Theorem XIV.3.1 [=]:

Proof of the Cayley-Hamilton Theorem.

This is a somewhat mysterious proof in any case, made more difficult to follow here by notational problems; but it has an intriguing quality once one has penetrated the obscurity. Let me try to show what is going on.

Let us assume without loss of generality that $E = k^{(n)}$, so that A can be regarded as a matrix. As in the preceding sections, we make E a $k[t]$-module by letting t act by A. We then form the matrix $B = tI_n - A \in \mathrm{Mat}_n(k[t])$, and let it act on $E^{(n)}$, i.e., on columns of elements of that module.

Now let us step back and note that, if we are stacking elements of E up in columns, then we should regard these elements as *row* vectors; thus elements of $E^{(n)}$ take the form of $n \times n$ matrices. Since t acts

on $E^{(n)}$ via an action on E, it acts on each row separately; so it should be regarded as *right* multiplication of these $n \times n$ matrices by A. Thus, if we left multiply one of these stacks of elements of E by $B = tI_n - A \in \mathrm{Mat}_n(k[t])$, the first term acts as right multiplication by A, while the second term, which does not involve t, but is a matrix of elements of k acting on our column, has the effect of *left* multiplication by A. Thus, B represents the difference between right and left multiplication by A. In general, this operator does not annihilate all of $E^{(n)}$; but it clearly annihilates the identity matrix; i.e., the stack whose ith term is the ith standard basis vector. Now if we can find a left multiple of B in $\mathrm{Mat}_n(k[t])$ that is a *scalar* matrix $c(t)I_n$ $(c \in k[t])$, then as a left multiple of B, it will still annihilate the identity matrix, hence the element $c(t)$ will annihilate all basis elements of E, hence will act trivially on E. Now in fact, the first line of Proposition XIII.4.16 tells us that for any matrix A over a commutative ring, the scalar matrix $\det(A)I_n$ is a left multiple of A; so the above argument shows that the operator $\det(B) = \det(tI_n - A) = P_A(t) \in \mathrm{Mat}_n(k[t])$ acts trivially on E, i.e., that $P_A(A) = 0$, as required.

There is another standard proof of this theorem, which uses ''brute force'' instead of ''magic'', but also involves a trick that is worth having under one's belt. Let me sketch it; you should not find it hard to fill in the details. (a) Verify that the theorem is true for a matrix A whose only nonzero entries lie in square diagonal blocks (as on p.555) *if* it is true for each of those blocks. (Use the fact that the characteristic polynomial of A is the product of the characteristic polynomials of those blocks.) (b) Verify that the theorem is true for a matrix of the form shown at the bottom of p.558, whose characteristic polynomial is $(t-\alpha)^e$. Deduce that it is true for any matrix in Jordan canonical form. (c) Verify that if it is true for a matrix A, it is true for any conjugate matrix CAC^{-1}, and deduce that it is true for any matrix over an algebraically closed field. (d) Since any integral domain embeds in a field, and any field embeds in an algebraically closed field, deduce that it is true for any matrix over an integral domain. (e) Since any commutative ring is a homomorphic image of an integral domain (for instance, a polynomial ring over $\mathbf{Z}$ in a sufficiently large set of indeterminates), deduce that it is true for any matrix over any commutative ring.

Note that steps (d) and (e) work because the Cayley-Hamilton Theorem for $n \times n$ matrices is equivalent to a set of identities for commutative rings (n^2 identities in n^2 variables. This is because the coefficients of the characteristic polynomial P_A of an $n \times n$ matrix A are polynomials in the entries of S, and the computation of $P_A(A)$ is in turn done by polynomial operations.) Clearly, if an identity holds in a ring, it holds in any subring and in any homomorphic image, and this is what we have used. □

P.562, Theorem XIV.3.2 [~]: Since the minimal polynomial of a linear transformation A is the generator of the ideal of polynomials that A satisfies, and the Cayley-Hamilton Theorem shows that the characteristic polynomial belongs to that ideal, the characteristic polynomial is divisible by the minimal polynomial. It is not hard to verify that the roots of the minimal polynomial are precisely the eigenvalues of A, hence these are also among the roots of the characteristic polynomial. Strikingly, Theorem XIV.3.2 shows that the eigenvalues of A are *all* the roots of its characteristic polynomial, and does so without using the Cayley-Hamilton Theorem.

In fact, none of the remaining results in this section use the Cayley-Hamilton Theorem in their proofs. Rather, the Cayley-Hamilton Theorem shows the importance of the characteristic polynomial of a matrix, and the remaining results give additional information about that important polynomial.

Incidentally, note that where we previously considered $tI_n - A$, we are now looking at $A - \lambda I_n$. The former order is more convenient in defining the characteristic polynomial, since it is nice to have that polynomial monic; the latter is slightly more natural when thinking about eigenvalues, since to say a vector x is in the kernel of this matrix is to say that ''A acts on x like multiplication by λ''. $A - \lambda I_n$ is also more convenient to write down in explicit cases, since one doesn't have to change the signs of all n^2 entries of A (and be plagued by computational errors if one forgets to change one of them).

P.562, Proof of Theorem XIV.3.2 [~]: Note the similarity of this proof to that of Theorem VI.4.1

(Artin's Theorem on Independence of Characters, p.283). In fact, not only the proofs but the results are related. For consider a k-vector-space E (not necessarily finite-dimensional) and a family $(A_i)_{i \in I}$ of linear maps $E \to E$. Suppose $w_1, \dots, w_m \in E$ are *simultaneous eigenvectors* of all the A_i (in general, with different eigenvalues for different A_i), and that for each $j \neq j'$, there is some A_i with respect to which the eigenvectors w_j and $w_{j'}$ have different eigenvalues. Then one can deduce from the present theorem (or more easily, adapt the proof of the theorem to show) that the vectors $w_1, \dots, w_m$ are linearly independent. I claim that this result, the linear independence of simultaneous eigenvectors, is equivalent to Artin's Theorem.

For let G be a monoid, and E the space of all k-valued functions on G. For each $g \in G$, let $A_g : E \to E$ denote the *shift operator* defined by $(A_g w)(h) = w(hg)$. It is not hard to show that an element $w \in E$ is a character if and only if w is an eigenvector of each of the operators A_g, and $w(1) = 1$. When this is so, the eigenvalue of the action of each A_g on w is $w(g)$, so distinct characters are associated with distinct families of eigenvalues. Hence by the result on linear independence of simultaneous eigenvectors, distinct characters are linearly independent.

Conversely, given a family of operators A_i on a vector space E, note that the monoid M of operators generated by the A_i acts on E, and each simultaneous eigenvector w of the A_i generates a 1-dimensional invariant subspace. The action of M on that subspace determines a character $M \to k$, and if we had a linear relation among eigenvectors inducing distinct characters, we could, by applying a linear functional, get a linear relation among these characters, a contradiction. So such eigenvectors are linearly independent.

P.564, Theorem XIV.3.7 [~]: This looks simpler if you only draw the exact sequence once, and express the condition that the squares in Lang's diagram commute as saying that the arrows are homomorphisms of $k[t]$-modules, where t acts on E', E and E'' by A', A and A'' respectively. Note that the theorem does not require k to be a field; but it is easiest to apply in that case, since the freeness hypothesis then holds automatically.

P.567, first two sentences of Proof of Theorem XIV.3.10 [>]: By ''the canonical decomposition in terms of matrices given in Theorem 2.4'' Lang means the Jordan canonical form. This gives an easy verification of the result when k is a field. Now, as in the alternative proof I sketched for the Cayley-Hamilton Theorem, the result for fields immediately implies the same result for integral domains, since every integral domain embeds in a field, and if the characteristic polynomial of M splits into linear factors over the integral domain, it certainly splits into the same factors over its field of fractions.

Now in the alternative proof of the Cayley-Hamilton Theorem, we passed from the case where k was an integral domain to the case where it was a general commutative ring using the observation that given any $n \times n$ matrix A over a commutative ring k, there existed an $n \times n$ matrix U over an integral domain R and a homomorphism $R \to k$ which carried the entries of U to the corresponding entries of A. To argue analogously here, we would need to know that given an $n \times n$ matrix A over a commutative ring k such that the characteristic polynomial of A is a product of linear factors, there exists an $n \times n$ matrix U over an integral domain R whose characteristic polynomial is also a product of linear factors, and a homomorphism $R \to k$ which carries U to A, and carries the factorization of the characteristic polynomial of U to the given factorization of the characteristic polynomial of A. (One would also need some elements to be mapped to the coefficients of the polynomial f of the statement of the theorem, but these cause no problem, so I will leave them out of the discussion here.)

Will there in general exist an integral domain R that can be so mapped? A natural way to attempt to get one would be to start with the polynomial ring in n^2 indeterminates, which is a commutative ring with a universal $n \times n$ matrix U, and then ''force'' the characteristic polynomial of U to factor, by adjoining to this ring n more indeterminates $y_1, \dots, y_n$ and imposing relations that equate the coefficients of the characteristic polynomial of U with the corresponding coefficients of $(t - y_1) \dots (t - y_n)$. Lang comments that the statement that this ring will be an integral domain ''is obvious to the reader who read the chapter on integral ring extensions'', and that ''the reader who has not can forget about this part of the

theorem'' (i.e., the case of the theorem where k is not an integral domain).

Fair enough. But thinking about this, I see a way of getting around the difficulty, which I will present below. The method is instructive, but since it represents a digression from the main subject, I may or may not, in a given year, wish to assign it as reading.

Alternative completion of the proof of Theorem XIV.3.10. (Optional.)

The complication in the above attempt to complete the proof arose in trying to weld together two universal constructions: the commutative ring with a universal $n \times n$ matrix, which is $\mathbf{Z}[x_{11}, x_{12}, \dots, x_{nn}]$, and the commutative ring with a universal degree-n polynomial that splits into linear factors $t - y_i$, which is $\mathbf{Z}[y_1, \dots, y_n]$. Each of these separately is a polynomial ring, hence an integral domain; but it is nontrivial to prove that the ring one gets by combining these constructions, to give the universal ring with an $n \times n$ matrix whose characteristic polynomial splits into linear factors, is also an integral domain. The trick below will be to use these constructions separately rather than together. We begin with the following result, which makes no mention of matrices.

Lemma XIV.3.c1. *Let n and d be positive integers. Then there exist polynomials $p_0, \dots, p_{n-1}$ in $n + d + 1$ indeterminates $W_0, \dots, W_{n-1}$, $Z_0, \dots, Z_d$ over $\mathbf{Z}$, with the property that if k is any commutative ring, $g(t) \in k[t]$ is any monic polynomial of degree n which factors into linear factors over k,*

$$g(t) \;=\; t^n + w_{n-1} t^{n-1} + \dots + w_0 \;=\; \textstyle\prod_i (t - s_i) \qquad (w_i,\, s_i \in k),$$

and $f(t) \in k[t]$ is any (not necessarily monic) polynomial of degree $\leq d$ over k,

$$f(t) \;=\; z_d t^d + z_{d-1} t^{d-1} + \dots + z_0,$$

then the monic polynomial $h(t)$ of degree n gotten by ''applying f to the roots of g'', i.e.,

$$h(t) \;=\; \textstyle\prod_i (t - f(s_i))$$

has the form

$$t^n \;+\; p_{n-1}(w_0, \dots, w_{n-1}, z_0, \dots, z_d)\, t^{n-1} \;+\; \dots \;+\; p_0(w_0, \dots, w_{n-1}, z_0, \dots, z_d).$$

Sketch of Proof. Form the polynomial ring in $n + d + 2$ indeterminates $S_1, \dots, S_n$, $Z_0, \dots, Z_d$, t over $\mathbf{Z}$, and in it define the polynomial

$$F(t) \;=\; Z_d t^d + Z_{d-1} t^{d-1} + \dots + Z_0,$$

and, using this, the polynomial

$$H(t) \;=\; \textstyle\prod_i (t - F(S_i)).$$

Clearly, $H(t)$ is invariant under all permutations of the variables S_i, hence each of its coefficients is symmetric as a polynomial in those variables; hence, by Theorem IV.6.1 (p.191), can be expressed in terms of the Z_i and the elementary symmetric functions in the S_i, the latter being precisely the coefficients of the polynomial

$$G(t) \;=\; \textstyle\prod_i (t - S_i).$$

Let $p_0(W_0, \dots, W_{n-1}, Z_0, \dots, Z_d)$, ..., $p_{n-1}(W_0, \dots, W_{n-1}, Z_0, \dots, Z_d)$ be the polynomials over $\mathbf{Z}$ such that on substituting the elementary symmetric functions in the S_i for the W's, one gets the coefficients of $H(t)$. It is easy to verify that these polynomials have the asserted property.

(You might wonder whether something stronger was needed where we called on Theorem IV.6.1, since we were talking about polynomials in $S_1, \dots, S_n$, Z_0, ..., Z_d that were only assumed symmetric in the first n indeterminates. But what we needed follows by applying that theorem to the coefficient in $\mathbf{Z}[S_1, \dots, S_n]$ of each monomial in Z_0, ..., Z_d.) □

To prove Lang's Theorem XIV.3.10, it will clearly suffice to show the stronger statement that given *any* $n \times n$ matrix M over a commutative ring k (not necessarily having a characteristic polynomial that splits into linear factors), and any polynomial f over k of degree $\leq d$, the coefficients of the characteristic polynomial of $f(M)$ are the values given by the above polynomials p_i on substituting for the W_i the coefficients of the characteristic polynomial of M, and for the Z_i, the coefficients of f. Since this assertion of equality is equivalent to a family of identities, it suffices to verify it in a universal situation, namely where M is a matrix of n^2 independent indeterminates, and the coefficients of f are $d+1$ independent indeterminates. Thus, we are reduced to the case where k is a polynomial ring over $\mathbf{Z}$. This is an integral domain; embedding it in its field of fractions, and that in its algebraic closure, we may factor the characteristic polynomial of M into linear factors, and apply the preceding lemma and the fact that Theorem XIV.3.10 holds over fields, to get the desired result.

Chapter XVI. The Tensor Product.

Re §XVI.1. Tensor product.

P.601, first paragraph [<]: Don't worry if you don't follow the latter half of this paragraph; Lang is talking about things that will be defined later on.

P.602, second paragraph [=]: Given a family $E_1, \ldots, E_n$, Lang talks here about forming a category whose objects are the n-linear maps on this family. In doing so, he is relying on the convention that a map determines its codomain (CAT 1, p.53). More intuitively, the objects of this category should be regarded as modules F *given with* multilinear maps of $E_1 \times \ldots \times E_n$ into them. (I would prefer actually to define the category that way. I made a similar remark concerning p.58, where Lang set up the category in terms of which he characterized the free abelian group on a set S.)

P.602, four lines below definition of **tensor product** [=]: After ''we know, of course, that a tensor product'', add ''if it exists''.

P.602, first two lines of next paragraph [=]: I recommend replacing the words ''generated by'' by the single word ''on'' in both lines. That is, we are not taking these n-tuples and using them to generate a submodule of an existing module, but taking the free module on this generating set.

P.604, two paragraphs beginning **Warning.** [>]. Here is a still more important warning: Not every member of a tensor product $E_1 \otimes \ldots \otimes E_n$ is of the form $x_1 \otimes \ldots \otimes x_n$. Elements of this form are called *decomposable tensors*, and they *generate* the tensor product module; consequently, one often proves that two homomorphisms on the tensor product are equal by verifying that they agree on all elements of that form. Also, the universal property of the tensor product tells us that one can construct a morphism out of the tensor product module by specifying its values on elements $x_1 \otimes \ldots \otimes x_n$, as long as the value specified is multilinear in $x_1, \ldots, x_n$. Seeing such arguments leads some students every year to imagine that the general element of $E_1 \otimes \ldots \otimes E_n$ *is* of the form $x_1 \otimes \ldots \otimes x_n$, and to give incorrect arguments based on this assumption.

As an example where such an argument leads to an incorrect conclusion, let $k \subseteq K$ be a field extension with $1 < [K:k] < \infty$, let us regard K as a k-module, and let us form the tensor product module $K \otimes_k K$. There is a natural k-module map from this module to K, given by $x \otimes y \mapsto xy$ (since the field multiplication of K is k-bilinear). Now it is easy to see that this map does not annihilate any nonzero element of the form $x \otimes y \in K \otimes K$; but if we concluded from this that it has trivial kernel, we would be quite wrong, since by Proposition XVI.2.3 in the next section, $K \otimes K$ has k-dimension $[K:k]^2 > [K:k]$.

P.604, Proposition XVI.1.1 [>]: In the proof, Lang will construct the map going in the opposite direction. But by the symmetry of the statement, this doesn't matter.

P.605, end of proof of Proposition XVI.1.1 [=]: Why is the map Lang has constructed an isomorphism? Because one can construct, in the same way, a map going the opposite way, and these two

maps take the generators $(x \otimes y) \otimes z$ and $x \otimes (y \otimes z)$ to one another. Hence their composites, in either order, are the identity.

Remark. When an interesting binary operation $*$ turns out to be associative, one generally finds that the function taking three elements x, y, z to the common value of $(x*y)*z$ and $x*(y*z)$ represents some elegant ternary analog of the given binary operation. (Example: look at the formula for the common value of the two sides in the proof of associativity of multiplication in a group ring.) So we should ask here what the common isomorphism class of $(E_1 \otimes E_2) \otimes E_3$ and $E_1 \otimes (E_2 \otimes E_3)$ represents. The answer is, of course, $E_1 \otimes E_2 \otimes E_3$. Thus, it would have been better to prove the proposition in the stronger form

$$(E_1 \otimes E_2) \otimes E_3 \;\cong\; E_1 \otimes E_2 \otimes E_3 \;\cong\; E_1 \otimes (E_2 \otimes E_3),$$

via isomorphisms characterized by the behavior

$$(x \otimes y) \otimes z \;\leftrightarrow\; x \otimes y \otimes z \;\leftrightarrow\; x \otimes (\otimes y \otimes z).$$

By symmetry, it suffices to show the existence of such an isomorphism between the center object and the left-hand object. The map going outward from the center object is very easy to construct: since $(x,y,z) \mapsto (x \otimes y) \otimes z$ is 3-linear, it induces a homomorphism $x \otimes y \otimes z \mapsto (x \otimes y) \otimes z$ by the universal property of $E_1 \otimes E_2 \otimes E_3$. One gets the map going the other way by the same method as in Lang's proof of the proposition: fixing one variable, proving bilinearity in the other two, using this to get a map on the 2-fold tensor product, etc..

Strictly speaking, even the result described above is just the simplest case of the result one should really prove, which is that given a finite family of finite families of modules $E_{1,1}, \dots, E_{1,n(1)}, \;\dots\; E_{m,1}, \dots, E_{m,n(m)}$, one has

$$\otimes_i (\otimes_j E_{ij}) \;\cong\; \otimes_{i,j} E_{ij}.$$

This is tedious to verify, but not really harder than the case sketched above.

P.607, end of §XVI.1 [>]. I will end the notes on this section with a brief digression, followed by a longer discussion of a topic to which I will return repeatedly in this chapter.

Digression: the duality game.

Suppose k is a field, and E and F are k-vector spaces, and we want to construct $E \otimes_k F$, but consider the system of generators and relations used on p.602 too ugly to dirty our hands with. Now though we can't easily capture $E \otimes_k F$ directly, it is easy to say what the space of linear functionals on that space should look like: By the universal property of $E \otimes_k F$ they will correspond to *bilinear forms* $E \times F \to k$. So the dual of the space we want can be identified with $L^2(E, F; k)$. Moreover, every vector space has a natural embedding in its double dual; so the space we want should embed in a natural way in $L^2(E, F; k)^\vee$. We find, in fact, that the image of $x \otimes y \in E \otimes_k F$ will be the functional $L^2(E, F; k) \to k$ that takes $\varphi \in L^2(E, F; k)$ to $\varphi(x, y) \in k$. Hence we may identify $E \otimes_k F$ with the subspace of $L^2(E, F; k)^\vee$ spanned by functionals of the form $\varphi \mapsto \varphi(x, y)$ $(x \in E,\ y \in F)$. (For a finite-dimensional vector space, the natural map into its double dual is an isomorphism, so if E and F are finite-dimensional, we can identify $E \otimes_k F$ with the whole space $L^2(E, F; k)^\vee$.)

You may have seen this given as the *definition* of a tensor product of vector spaces in some other course. But though the construction is ingenious, and worth seeing, I think it really misses the point of what tensor products are about. It is also inapplicable to general modules over a ring that is not a field.

A still more primitive way of constructing the tensor product of two vector spaces E and F is to take bases $(\xi_i)_{i \in I}$ and $(\eta_j)_{j \in J}$ of these spaces, and define $E \otimes_k F$ to be a vector space having for basis a set of symbols $(\xi_i \otimes \eta_j)_{i \in I, j \in J}$. This is very concrete, but does not make it clear how the spaces one gets using different choices of bases are ''naturally isomorphic''. (They are isomorphic because they have the same dimension, but one wants to specify the isomorphism that for every $x \in E,\ y \in F$ sends the

element named $x \otimes y$ in one space to the element named $x \otimes y$ in the other.) It is also, like the previous construction, inapplicable to general modules over rings that are not fields.

I believe that this was the earliest definition of tensor product; the definition in terms of the dual of the space of bilinear forms was seen as an advance over this, because it gave a ''basis-free'' definition. But I think the generators-and-relations construction is preferable to both of these.

Tensor products in the noncommutative context.

If we want to define a concept of tensor product of modules over a not necessarily commutative ring S, we need an appropriate version of the concept of bilinear map. This is given in the following definition (though as I will mention later, there is not unanimity as to the language used).

Definition XVI.1.c1. *Let S be a (not necessarily commutative) ring, E a right S-module, F a left S-module, and A an abelian group. Then a map $\beta: E \times F \to A$ is called* bilinear *if it satisfies*

(c43) $$\beta(x + x', y) = \beta(x,y) + \beta(x',y) \qquad (x, x' \in E,\ y \in F),$$

(c44) $$\beta(x, y + y') = \beta(x,y) + \beta(x,y') \qquad (x \in E,\ y, y' \in F),$$

(c45) $$\beta(xs, y) = \beta(x, sy) \qquad (x \in E,\ s \in S,\ y \in F).$$

(Examples: If S is any ring, then we may regard S as a right S-module, a left S-module, and an abelian group, and the multiplication operation $S \times S \to S$ clearly satisfies the above definition. If k is any ring and m, n, p are positive integers, then in the matrix notation I introduced in the comments to §XIII.1, we may regard ${}^m k^n$ as a right ${}^n k^n$-module, ${}^n k^p$ as a left ${}^n k^n$-module, and ${}^m k^p$ as an abelian group. Then the matrix multiplication operation ${}^m k^n \times {}^n k^p \to {}^m k^p$ is bilinear with respect to these structures.)

We can now construct from any E and F as in the above definition an abelian group A with a universal map having the above properties. The construction is exactly like that on p.602, mutatis mutandis, so I will leave the details to you, and simply make

Definition XVI.1.c2. *Let S, E and F be as in the preceding definition. Then $E \otimes_S F$ will denote the abelian group having a universal bilinear map $\otimes: E \times F \to E \otimes_S F$ (constructed using generators $x \otimes y$ for all $x \in E$, $y \in F$, and relations corresponding to* (c43)-(c45) *above*).

Note that the tensor product construction is a *functor* in each variable. I.e., given a right S-module homomorphism $f: E \to E'$, we get a homomorphism of abelian groups $E \otimes F \to E' \otimes F$ which acts by $x \otimes y \mapsto f(x) \otimes y$, and this construction respects composition and identity homomorphisms; and the same is true for maps of the second argument. These constructions can also be seen to respect *addition* of module homomorphisms. Finally, given both a homomorphism $f: E \to E'$ and a homomorphism $g: F \to F'$, the diagram of induced maps

$$\begin{array}{ccc} E \otimes F & \to & E' \otimes F \\ \downarrow & & \downarrow \\ E \otimes F' & \to & E' \otimes F' \end{array}$$

commutes, since both composite maps take $x \otimes y$ to $f(x) \otimes g(y)$. These facts will soon be important to us.

It may seem disappointing that the definition of bilinear map only involves an abelian group structure on A, and hence that our tensor product object is likewise only an abelian group. We might hope that if in the context of Definition XVI.1.c1 we also had, say, a right S-module structure on A, there would be a natural condition relating this to the right S-module structure on E, which we could expect to hold under favorable circumstances; for instance, the condition $\beta(xs, y) = \beta(x, y)s$. But in fact, this condition turns

out to be unnatural; for if it holds, we see that for any $s_1, s_2 \in S$ we have

$$\beta(xs_1s_2, y) \;=\; \beta(xs_1, s_2y) \;=\; \beta(x, s_2y)s_1 \;=\; \beta(xs_2, y)s_1 \;=\; \beta(xs_2s_1, y).$$

Thus, $\beta(x(s_1s_2 - s_2s_1), y) = 0$, so that β involves E only via the factor-module $E/E[S,S]$, which is a module over the commutative ring $S/[S,S]$. The same would hold by similar arguments for the left module structure of F, and for the right S-module structure of the submodule of A generated by the image of β. Thus we would really be looking at a bilinear map of modules over the commutative ring $S/[S,S]$, and this would exclude all essentially noncommutative examples, such as those indicated above.

What was wrong here, however, was only that we were trying to imitate too slavishly the commutative situation! To be able to state the correct extension of the concept of bilinear map, we first need

Definition XVI.1.c3. *Let R and S be rings. Then an* (R,S)-bimodule *will mean an abelian group A given with both an action of R by endomorphisms on the left* (*a left R-module structure*) *and an action of S by endomorphisms on the right* (*a right S-module structure*), *related by the identity*

(c46) $$(rx)s \;=\; r(xs) \qquad (r \in R,\ x \in A,\ s \in S),$$

i.e., such that R acts by endomorphisms of the S-module structure, equivalently, such that S acts by endomorphisms of the R-module structure.

Examples: Any ring S is an (S,S)-bimodule, via its internal multiplication. We can get further interesting examples by starting with examples of the above sort, taking two subrings R and T of S, and noting that S will then also be an (R,T)-bimodule, and may have interesting (R,T)-subbimodules.

For example, consider the division ring $\mathbf{H}$ of quaternions (Exercise XIII:L7, p.545) as a $(\mathbf{C},\mathbf{C})$-bimodule. One sees that it is the direct sum of two sub-bimodules, namely $\mathbf{C}$ itself, and $\mathbf{C}j$. Each of these is 1-dimensional as a right and as a left $\mathbf{C}$-module, but they have nonisomorphic bimodule structures, since right and left multiplication by $i \in \mathbf{C}$ agree on one, but not on the other.

If we take for S the *Weyl algebra*, generated over $\mathbf{R}$ by elements X and Y satisfying the relation $XY - YX = 1$ (corresponding to the differential operators of multiplication by x and differentiation with respect to x on polynomial functions on the real line; see this Companion, notes to p.85), then we can regard this as an $(\mathbf{R}[X], \mathbf{R}[X])$-bimodule. You may find it interesting to examine the structure of the $(\mathbf{R}[X], \mathbf{R}[X])$-subbimodule generated by Y (whose general element can be written $Yf(X) + g(X)$). These two examples both show that even a bimodule over a commutative ring can have more structure than just that of a ''module'' in the sense of commutative ring theory.

Finally, as I mentioned in the notes to §XIII.1 (though I had not yet formally defined the concept of bimodule), a matrix ring ${}^mk^n$ is an $({}^mk^m, {}^nk^n)$-bimodule.

Note that in the examples of *bilinear maps* given after Definition XVI.1.c1, the objects E, F and A actually had natural bimodule structures, and the bilinear maps ''respected'' these. The sense in which this held is formalized in

Definition XVI.1.c4. *Suppose that in the situation of Definition XVI.I.c1, E is in fact an (R,S)-bimodule, and A a left R-module, for some ring R. Then we will call β bilinear as a map from the (R,S)-bimodule E and the left S-module F to the left R-module A if in addition to* (c43)-(c45), *it satisfies*

(c47) $$\beta(rx, y) \;=\; r\beta(x,y) \qquad (r \in R,\ x \in E,\ y \in F).$$

If, instead, F is an (S,T)-bimodule and A a right T-module for some ring T, then we will call β bilinear as a map from the right S-module E and the (S,T)-bimodule F to the right T-module A if, in addition to (c43)-(c45), *it satisfies*

(c48) $$\beta(x, yt) = \beta(x,y)t \qquad (x \in E,\ y \in F,\ t \in T).$$

Finally, if E *is an* (R,S)*-bimodule,* F *an* (S,T)*-bimodule, and* A *an* (R,T)*-bimodule, then we shall say that* β *is bilinear with respect to these bimodule structures if* (c47) *and* (c48) *both hold.*

This would seem to call for further universal constructions, e.g., ''the tensor product of an (R,S)-bimodule E with a left S-module F'', which would be a left R-module, and so forth. But marvelously, no new constructions are needed. For let E be an (R,S)-bimodule and F a left S-module, and let us form their tensor product $E \otimes_S F$ as a right and a left S-module, as before. Then since R acts on E by endomorphisms of its right S-module structure, and the tensor product construction is functorial, each element of R induces an endomorphism of the abelian group $E \otimes_S F$. It is easy to verify (from the observations we made earlier concerning the functoriality of $\otimes$) that these endomorphisms constitute a left R-module structure on $E \otimes_S F$, with respect to which the map $(x,y) \mapsto x \otimes y$ satisfies (c47), and the universal property of that map among maps bilinear for the S-module structures on E and F immediately gives the universal property for maps that are bilinear in the stronger sense. Likewise, if, instead, F has an (S,T)-bimodule structure, then $E \otimes_S F$ acquires a right T-module structure making $\otimes$ universal in the appropriate strengthened sense, and if both E and F have bimodule structures as above, then $E \otimes_S F$ becomes an (R,T)-bimodule, and $\otimes$ becomes a universal bilinear map of bimodules.

It may still seem awkward to have four different cases to consider. Fortunately, even this is only apparent. Just as an abelian group is equivalent to a $\mathbf{Z}$-module, so a right S-module is equivalent to a $(\mathbf{Z},S)$-bimodule, a left S-module is equivalent to an $(S,\mathbf{Z})$-bimodule, and an abelian group is equivalent to a $(\mathbf{Z},\mathbf{Z})$-bimodule, in the sense that any structure of the indicated ''weaker'' sort extends *uniquely* to a structure of the ''stronger'' sort, and any homomorphism of the ''weaker'' structure is in fact a homomorphism of the ''stronger'' structure. So all of the concepts of bilinearity introduced in this section can be subsumed in the concept of a bilinear map from an (R,S)-bimodule cross an (S,T)-bimodule to an (R,T)-bimodule; and all our versions of the tensor product can be subsumed in the statement that the tensor product with respect to S-module structures of an (R,S)-bimodule and an (S,T)-bimodule has a natural structure of (R,T)-bimodule, with respect to which the map $(x,y) \mapsto x \otimes y$ is bilinear in our strongest sense, and is universal for that property. Of course, we will continue to refer to left and right modules and to abelian groups; we will simply know that any statement we make about bimodules applies to these as special cases.

A module E over a *commutative* ring k can be regarded either as a left k-module, as a right k-module, or as a (k,k)-bimodule in which $xc = cx$ for all $x \in E$ and $c \in k$. We find that the concept of bilinear map $E \times F \to A$ of k-modules as defined in Lang for modules over commutative rings agrees with our above definition of bilinear map when the ''bimodule'' interpretations are used, and that the concept of tensor product, which we have seen in our context to depend essentially only on a right module structure on E and a left module structure on F, also agrees with the definition for modules over commutative rings whenever k-modules E and F are interpreted in terms of at least this much structure.

I should mention here a point of language that I alluded to earlier. Some writers feel that since ''bilinear'' means ''twice linear'', the condition that a map from an (R,S)-bimodule cross an (S,T)-bimodule to an (R,T)-bimodule be ''bilinear'' should only mean that it is R-linear in the first argument and T-linear in the second argument, i.e., that it satisfies (c43), (c44), (c47) and (c48). Though these authors agree that (c45) is equally important, they feel compelled to give it a different name, calling a map that satisfies it ''balanced'' or ''middle-linear''. So they would refer to as a ''bilinear and balanced'' or ''bilinear and middle-linear'' map what I call a ''bilinear map''. I will continue to use the term ''bilinear'' to include condition (c45).

An easy exercise, which you should think through, is to verify that the object having the universal property of a *free* (R,S)*-bimodule on one generator* is $R \otimes_{\mathbf{Z}} S$, the generator being $1 \otimes 1$.

Though it would not have been apparent from Definition XVI.I.c1 that there was any way to generalize the concept of bilinear map to get concepts of n-linear maps $E_1 \times \ldots \times E_n \to F$ for higher n, we now see

that these will make sense if for rings $R_0, \dots, R_n$, each E_i is an (R_{i-1}, R_i)-bimodule and F is an (R_0, R_n)-bimodule. And (just as for modules over commutative rings) one can define the n-fold tensor product of such a family, and verify that it can be obtained by iterating the 2-fold tensor product construction. In fact, everything Lang does in the above section goes over to this context, except Proposition XVI.1.2 (commutativity of the tensor product construction), which is meaningless in the general context where E is an (R,S)-bimodule and F an (S,T)-bimodule, and still false (Exercise XVI.I:5) even if we happen to have $R = S = T$ so that the two sides could be compared.

In my comments on the remaining sections of this chapter of Lang, I will often note that definitions and results that he gives for tensor products $E \otimes_R F$ of modules over a commutative ring R go over to the context of tensor products $E \otimes_R F$ of a right R-module E and a left R-module F, for a not necessarily commutative ring R. I will seldom refer to bimodules, but I hope that the above discussion of these has shown that tensor products of the sort just described are the key to a wider theory. I will generally call the base-ring R, as Lang does; it was only the desire to put it in the middle of a 3-letter sequence that led me to call it S in most of the above development.

Incidentally, if we assume a commutative ring k to be given, then throughout the above discussion we can replace ''ring'' by ''k-algebra'', ''abelian group'' by ''k-module'', and ''bimodule'' by ''k-centralizing bimodule'', i.e., bimodule such that the actions of elements of k on the left and on the right agree. It is easy to show that if R, S, T are k-algebras, and E and F are respectively a k-centralizing (R,S)-bimodule and a k-centralizing (S,T)-bimodule, then the tensor product $E \otimes_S F$ constructed as above will be a k-centralizing (R,T)-bimodule, and will be universal among k-centralizing (R,T)-bimodules given with bilinear maps of $E \times F$ into them. But for simplicity, I will not speak of this again.

Re §XVI.2. Basic properties.

P.607, boxed display [=]: The isomorphism between the left-hand and right-hand sides can be expressed as saying that for a fixed R-module F, the functor $E \mapsto E \otimes F$ is *left adjoint* to the functor $G \mapsto L(F,G)$, or in more compact notation, $- \otimes F$ is left adjoint to $L(F,-)$. By interchanging the roles of E and F, we see that $E \otimes -$ is also left adjoint to $L(E,-)$.

If you have had a bit more category theory than Lang gives, you may know that every left adjoint functor respects colimits, in particular, coproducts and direct limits. The fact that $E \otimes -$ respects direct limits was noted at the top of p.604; the fact that it respects coproducts (which in categories of modules are direct sums) will be noted on p.608. The explicit verifications given are unnecessary if one knows the general category-theoretic fact.

What form does this boxed display take in the noncommutative context? There are several versions of the statement; one good enough for our present purposes is to take E to be a right R-module, F to be a left R-module, and G to be an abelian group. It turns out that F being a left R-module makes the set $L(F,G)$ of abelian group homomorphisms $F \to G$ into a *right* R-module, by the definition $(fr)x = f(rx)$ ($f \in L(F,G)$, $r \in R$, $x \in F$). (Have I violated here my rule of writing homomorphisms of left modules on the right? No, because though F is a left R-module, f is not a left R-module homomorphism. In fact, if we regard F as an $(R, \mathbf{Z})$-bimodule, we can describe f as a homomorphism of right $\mathbf{Z}$-modules, and fr as the result of composing it on the right with the $\mathbf{Z}$-module endomorphism r of F.) If we then let $L(E,\ L(F,G))$ denote the abelian group of module homomorphisms between these *right* R-modules, the boxed result is easily established as a functorial isomorphism of abelian groups. In particular, $- \otimes F$ is still a left adjoint functor.

P.609, Proposition XVI.2.3 [~]: This says that if E is free on a basis indexed by a set I, then $E \otimes F$ is isomorphic to the direct sum of an I-tuple of copies of F.

In the noncommutative context, one can say that if E is free as a right R-module on a basis indexed by I, then as an *abelian group*, $E \otimes F$ is isomorphic to the direct sum of an I-tuple of copies of F. If F is in fact an (R,S)-bimodule, then this is an isomorphism of right S-modules.

The corollary to this result can likewise be deduced in the noncommutative context, in the form ''If a right R-module E is free on a basis indexed by I, and an (R,S)-bimodule F is free as a right S-module on a basis indexed by J, then $E \otimes_R F$ is free as a right S-module, on the obvious set indexed by $I \times J$''. Note the curious fact that nothing is assumed about the left R-module structure on F, yet it has to be there for the result to make sense.

As usual, each of these statement also holds with the roles of right and left reversed; e.g., if we tensor a right R-module E with a free left R-module F, the result has the abelian group structure of a direct sum of copies of E.

P.610, Proposition XVI.2.5 [>]: This is the one result of this section for which there is no evident noncommutative version. (If E is a right R-module, that does not create any sort of module structure on $\mathrm{End}_R(E)$, so it is not clear over what one could take such a tensor product.)

Note that the homomorphism displayed in this proposition will exist without the assumption that E and F are free of finite rank, but in such cases it need not be one-to-one or onto. Some examples where it is not, and an additional class of cases where it *is* an isomorphism, are given in Exercise XVI.2:1. Incidentally, in the context of the material we have had so far, this map can only be considered a homomorphism of modules, but §XVI.6 will show how to make a tensor product of R-algebras an R-algebra, and in that context it will be clear that the map is an algebra homomorphism.

P.612, lines 3-5 [>]: Why must the above sequence be exact if the original sequence splits? Because a split short exact sequence $0 \to E' \to E' \oplus E'' \to E'' \to 0$ is a termwise direct sum of two short exact sequences $0 \to E' \xrightarrow{\mathrm{id}} E' \to 0 \to 0$ and $0 \to 0 \to E'' \xrightarrow{\mathrm{id}} E'' \to 0$. Each of these clearly remains of the same form under tensoring with F, hence remains exact; and since tensoring with F is a left adjoint functor, it preserves direct sums. Hence it takes the given sequence to the direct sum of two exact sequences, which is an exact sequence.

P.612, Proposition XVI.2.7 [>]: This result does not need a separate proof – we can get it by tensoring the short exact sequence $0 \to \mathbf{a} \to R \to R/\mathbf{a} \to 0$ with E and applying Proposition XVI.2.6 (with the roles of first and second tensor factor reversed). Indeed, this gives an exact sequence $\mathbf{a} \otimes E \to R \otimes E \to (R/\mathbf{a}) \otimes E \to 0$; the middle term can be identified with E, and the image of the left-most term in that term is $\mathbf{a}E$, so the rightmost term is isomorphic to $E/\mathbf{a}E$, as claimed.

In the noncommutative context, the result makes sense for E a left R-module and $\mathbf{a}$ a right ideal; in that case, $E/\mathbf{a}E$ is an abelian group. If $\mathbf{a}$ is in fact a 2-sided ideal, i.e., a sub-(R,R)-bimodule of R, then the isomorphism described will be an isomorphism of left R-modules (or left $(R/\mathbf{a})$-modules).

P.612, end of §XVI.2 [>]. We have seen that the functor $F \otimes -$ takes short exact sequences to sequences that are *almost* short exact, with possible failure of exactness at the left end. Can we find an example where this failure actually occurs? As Lang observed earlier on this page, in such an example the original sequence must be non-split; so we first need an example of a non-split short exact sequence of modules. We know that the rightmost module of such a sequence must be non-projective; in particular, non-free. The simplest examples of non-free modules are the $\mathbf{Z}$-modules $\mathbf{Z}/p\mathbf{Z}$. There are two easy examples of nonsplit short exact sequences with this module as their rightmost term:

$$0 \to \mathbf{Z} \xrightarrow{p} \mathbf{Z} \to \mathbf{Z}/p\mathbf{Z} \to 0, \tag{c49}$$

$$0 \to \mathbf{Z}/p\mathbf{Z} \xrightarrow{p} \mathbf{Z}/p^2\mathbf{Z} \to \mathbf{Z}/p\mathbf{Z} \to 0. \tag{c50}$$

What shall we tensor these with? Not a free module; tensoring with a free module sends any short exact sequence to a direct sum of copies of itself. The non-free $\mathbf{Z}$-module that suggests itself in relation to the above two sequences is obviously $\mathbf{Z}/p\mathbf{Z}$. And we in fact find that on tensoring either of them with this module (and dropping the initial 0, so as not to write down a sequence that looks like a short exact sequence but isn't), we get

$$\mathbf{Z}/p\mathbf{Z} \xrightarrow{0} \mathbf{Z}/p\mathbf{Z} \xrightarrow{\mathrm{id}} \mathbf{Z}/p\mathbf{Z} \to 0,$$

in which the left-hand arrow is indeed not one-to-one.

The examples (c49) and (c50) are useful test-cases whenever a non-split exact sequence is needed.

Re §XVI.3. Flat modules.

Pp.612-613, conditions **F1** to **F3** [>]: These are all versions of the condition ''$F \otimes -$ preserves exactness''. It is useful to know the equivalence of strong and weak forms of a condition, so that, for instance, when one has to prove a module flat, one only needs to establish the weak statement **F3**, while when flatness is known, one has available the strong statement **F1**, which, applied simultaneously to each step in an exact sequence, says that $F \otimes -$ takes arbitrary exact sequences to exact sequences. (The significance of **F2** is indicated in Exercise XVI:3.1.) You should think through the proof of equivalence that Lang sketches, noting that it is valid in the context where F is a right module and the E's are left modules. Most of the other results proved in this section will also be true in the not-necessarily-commutative context, so let us make the convention that as far as possible, the modules which Lang tensors with exact sequences will be taken to be right modules in our discussion, and those comprising the exact sequences will be taken to be left modules. (The corresponding results with right and left modules interchanged will of course also be true, but we will not refer to them. Unfortunately, Lang vacillates between writing the modules for which he is considering the flatness property on the right and on the left of the tensor product sign, so that about half the time you will have to implicitly reverse his order to follow our convention of discussing flat *right* modules.)

P.613, Proposition XVI.3.1 [>]: Statement (i) is immediate, since $R \otimes -$ is isomorphic to the identity functor. Statement (ii) is a consequence of the facts that $(\oplus F_i) \otimes E \cong \oplus (F_i \otimes E)$, and that a direct sum of sequences is exact if and only if all summands are exact. From (i) and (ii) we see that a direct sum of copies of R, i.e., any free module, is flat, hence a direct summand in a free module, i.e., a projective module, is flat, which is (iii).

Tensor products also respect direct limits (p.604, top, also p.639, Exercise XVI:L12), and direct limits of exact sequences are exact (immediate), hence we also have

Lemma XVI.3.c1. *A direct limit of flat modules is flat.* □

The facts that tensor products respect direct sums and direct limits, used in proving the results we have just been discussing, are instances of the general fact mentioned earlier that tensor products respect colimits. However, most sorts of colimits do not respect exactness, and so do not carry flat modules to flat modules. For instance, the *cokernel* of a module-map $f: F \to F'$ can be characterized as the colimit of a diagram based on a parallel pair of maps, f and the zero homomorphism; but $\mathbf{Z}/p\mathbf{Z}$, which is the cokernel of the map $\mathbf{Z} \xrightarrow{p} \mathbf{Z}$ of flat $\mathbf{Z}$-modules, is not flat, since we saw that (c49) and (c50) do not remain exact under tensoring with it. In this connection you may find it instructive to write down the case of the snake lemma (pp.158-159) in which the middle horizontal sequences are two copies either of (c49) or of (c50), and the connecting vertical arrows are multiplication by p.

Let us now turn to some important

Further criteria for flatness

There are a couple of characterizations of flat modules rather different from **F1**-**F3**, which Lang relegates to an exercise (XVI:L13). Since it will be useful to refer to these in discussing the results to come, but the exercise is a hard one and the hint Lang gives for a key step is inadequate, let me develop the result here.

From Lemma XVI.3.c1 above, we see in particular that *a direct limit of free modules of finite rank is flat*. Our main result below will be the converse. In getting this, we shall use a third equivalent condition, which I will now motivate.

Given a *relation* holding among elements $x_1, \dots, x_n$ in any sort of algebraic structure, let us say (in this discussion) that the relation ''holds for good reason'' if there exist elements $y_1, \dots, y_m$ such that the x's can be expressed by certain formulas in the y's, in such a way that any n-tuple of elements of an algebraic structure of the given type that can be expressed by those formulas in terms of an m-tuple of

elements must satisfy the given relation. As an example of what I mean, suppose two elements x_1, x_2 of a group G commute with each other. In any group, any two powers of an element commute, so if we can find an element $y \in G$ and integers s and t such that $x_1 = y^s$ and $x_2 = y^t$, then x_1 and x_2 commute ''for good reason''. (In contrast, you will not be able to find any such ''good reason'' for the elements (1,0) and (0,1) of $\mathbf{Z} \times \mathbf{Z}$ to commute – if we regard $\mathbf{Z} \times \mathbf{Z}$ simply as a group. If we regard it as an *abelian* group, then of course any pair of its elements commutes ''for good reason''.) For another example, consider a pair of elements x_1, x_2 of an abelian group A which satisfy $4x_1 = 6x_2$. If we can write $x_1 = 3y$, $x_2 = 2y$, then they satisfy the above relation for good reason; and in fact, if A is torsion-free, you will find that taking $y = x_1 - x_2$, we will indeed have $x_1 = 3y$ and $x_2 = 2y$; on the other hand, if $A = \mathbf{Z}/2\mathbf{Z}$ and $x_1 = x_2 = 1$, there will be no y with that property.

Rather than going further into these examples (and providing examples where the number of y's used must be greater than 1), let us note how to make the above loosely stated condition precise. I spoke of a relation holding among a family of elements $x_1, \dots, x_n$ that follows from their being expressible by certain formulas in terms of another family $y_1, \dots, y_m$. A precise way to say this is that there exist elements $X_1, \dots, X_n$ in the *free* object (free group, free abelian group, free R-module, or whatever) F on m generators $Y_1, \dots, Y_m$, such that $X_1, \dots, X_n \in F$ satisfy the given relation, and such that there exists a homomorphism from F into our given object which carries $X_1, \dots, X_n$ to $x_1, \dots, x_n$.

Let us at this point restrict attention to the specific sort of object we are interested in here, right modules over a fixed ring R. Suppose elements $x_1, \dots, x_n$ of a right R-module E satisfy a relation $\Sigma_j \; x_j a_j = 0$. This will hold for good reason if and only if there exist an integer m, elements $y_1, \dots, y_m \in E$, and equations $x_j = \Sigma \; y_i b_{ij}$ holding in E, such that if, in the corresponding free right module ${}^m R$, we form the corresponding linear combination of the free generators e_i, namely the columns b_j of the matrix $B = (b_{ij})$, then these will also satisfy the relation $\Sigma \; b_j a_j = 0$. If we write A for the column vector (a_j), then the latter relation becomes the matrix equation $BA = 0$.

Lemma XVI.3.10 (which you should read now – although it is on p.621, it does not call on any of the results proved in the intermediate pages) thus says precisely that *any* relation holding among elements of a *flat* module holds ''for good reason'', in this sense! (Actually, it says this for a finitely generated flat module, but finite generation is never used, and should be dropped from the statement.) Note that, unfortunately, to maintain our practice of making the module we are testing for flatness a right module, we must reverse the orders of all multiplications in the proof, both of elements and of tensor factors. In the discussion below I will also, for convenience, take the subscripts i and j to be interchanged relative to the way Lang uses them in that proof, and, for compatibility with Exercise XVI:L13, rename M as E. Note that since we are taking E to be a right module, the first displayed sequence in the proof should be regarded as an exact sequence of *left* R-modules,

$$0 \;\to\; K \;\to R^n \xrightarrow{\;A\;} R.$$

Assuming you have gone through the proof of the lemma, I now claim a little more than the lemma asserts; namely, that given elements $x_1, \dots, x_n \in E$ as above, and any *finite family* of relations that they satisfy,

$$\Sigma \, x_j a_{jk} \;=\; 0 \qquad (k = 1, \dots, r), \tag{c51}$$

these relations must *simultaneously* hold for good reason; that is, it is possible to write the x_j as linear combinations of other elements y_i in such a way that the corresponding elements X_j of the free module on generators Y_i satisfy all r of these relations.

On the one hand, this result could be proved inductively from the lemma as stated. For given a family of r relations (c51) satisfied by the x_j, the lemma allows us to express the x_j in terms of other elements y_i so that *one* of these relations becomes a consequence of this way of expressing them. The remaining $r-1$ relations, on substituting for the x's these expression in the y's, translate to $r-1$ relations satisfied by the y's. We now use the lemma again, to express the y's in terms of elements z_h so that one

more relation is eliminated, and so on.

However we can, more briefly, get this result by going through the proof of Lemma XIV.3.10 with the $n\times 1$ matrix (a_j) replaced by the $n\times r$ matrix (a_{jk}); in particular, with the first display replaced by

$$0 \to K \to R^n \xrightarrow{A} R^r.$$

Let us now make one further restatement of the above result. To choose a family of elements $x_1, \dots, x_n$ of a right module E satisfying a family of r relations $\Sigma_j\, x_j a_{jk} = 0$ is equivalent to choosing a homomorphism into E from the right module P presented by n generators $X_1, \dots, X_n$ and r relations $\Sigma_j\, X_j a_{jk} = 0$, i.e., the right module with presentation

$$^rR \xrightarrow{A} {}^nR \to P \to 0.$$

Our result that every family of elements satisfying such a system of relations is the image of a family of elements in a free module which satisfies the same system of relations thus translates to the statement that every homomorphism from a *finitely presented* right R-module P into E factors through a free module F (without loss of generality, of finite rank).

Turning to Exercise XVI:L13 on p.639, we see that we have proved (i) $\Rightarrow$ (ii) $\Rightarrow$ (*) $\Leftrightarrow$ (iv), where (*) is the result of Lemma XIV.3.10, generalized to finite families of relations (c51).

(We have sidestepped condition (iii) of the exercise. If you want to work that condition in, note that in our context, the displayed homomorphism thereof should be written

$$E \otimes_R \mathrm{Hom}(P, R) \to \mathrm{Hom}(P, E),$$

and that the condition that it be surjective is equivalent to saying that every homomorphism $P \to E$ is a sum of finitely many homomorphisms that each factor through R. But such a finite sum is equivalent to a single homomorphism that factors through a direct sum of copies of R, and the condition, so translated, is easily seen to be equivalent to (iv).)

Before approaching the final implication, (iv)$\Rightarrow$(i), let us note that any module E is a direct limit of finitely presented modules. Indeed, take any presentation of E by a generating set S (e.g., all elements of E) and a set T of relations satisfied by the elements of S in E (e.g., all such relations). Now consider the set I of all pairs (S_0, T_0) such that S_0 is a finite subset of S, and T_0 is a finite subset of T consisting of relations which involve only elements of S_0. We may partially order I by componentwise inclusion, and we see that as such it is *directed* (every pair of elements has a common upper bound). Now for every $(S_0, T_0) \in I$, let us define $P_{(S_0, T_0)}$ to be the module presented by generators S_0 and relations T_0. Whenever $(S_0, T_0) \leq (S_1, T_1)$ in I, we clearly get a unique module homomorphism $P_{(S_0, T_0)} \to P_{(S_1, T_1)}$ taking the image in $P_{(S_0, T_0)}$ of each element of S_0 to the image in $P_{(S_1, T_1)}$ of the same element, and these maps make the $P_{(S_0, T_0)}$ a directed system of modules. These modules all have natural maps into E which make commuting triangles with the above maps, and the universal property of E as the module presented by the given generators and relations immediately implies its universal property as a direct limit of these finitely presented modules.

We now seem very close to having the implication (iv)$\Rightarrow$(i). Given a module E that satisfies (iv), let us express it as $\varinjlim_I P_i$ for some directed system $(P_i)_{i\in I}$ of finitely presented modules, denoting the connecting homomorphisms by

$$p_{j,i} : P_i \to P_j \qquad (i < j \text{ in } I),$$

and the limit homomorphisms by

$$p_i : P_i \to E.$$

We may use (iv) to factor each of the p_i through a free module of finite rank:

(c52) $$P_i \xrightarrow{h_i} F_i \xrightarrow{f_i} E.$$

But here we run into a problem. Condition (iv) is not a functorial statement – it says that for *each* map from a finitely presented module P_i into E we can get a factorization through an F_i, but not that we can do this so that for each map between such modules $P_i \to P_j$ ''over'' E (i.e., making a commuting triangle with the maps to E) we can get a map $F_i \to F_j$ ''over'' E, and certainly not that we can do this in a way that respects composition on our infinite family of modules and maps.

Indeed, once the factorizations (c52) have been chosen, it may not be possible to get any corresponding maps among the F's. For given $i<j$, note that $p_i(P_i) = p_j p_{j,i}(P_i) \subseteq p_j(P_j)$; but F_i and F_j may have been chosen so that $f_i(F_i)$ is not similarly contained in $f_j(F_j)$; in which case we cannot get a map $F_i \to F_j$ over E.

I claim, however that (assuming the factorizations (c52) chosen in an arbitrary but fixed way), if we take j ''far enough above i'', this problem vanishes. Precisely, for each $i \in I$, the facts that E is the direct limit of the P_j and that $f_i(F_i)$ is finitely generated tell us that there will exist some $i' > i$ such that $f_i(F_i) \subseteq p_{i'}(P_{i'})$. Then for any $j \geq i'$, we will have $f_j(F_j) \supseteq p_j(P_j) \supseteq p_{i'}(P_{i'}) \supseteq f_i(F_i)$. Since F_i is free, this allows us to map F_i to any such F_j over E; in particular, to $F_{i'}$. In fact, we will find it useful to do something stronger, which the condition $f_i(F_i) \subseteq p_{i'}(P_{i'})$ also shows is possible, namely, to choose a homomorphism $\theta_i : F_i \to P_{i'}$ over E.

Unfortunately, the map θ_i may still not be very ''good'', in the sense that the composite $P_i \xrightarrow{h_i} F_i \xrightarrow{\theta_i} P_{i'}$ may not agree with the map $p_{i',i} : P_i \to P_{i'}$ of the original directed system. However, note that for every $x \in P_i$, the two elements $\theta_i h_i(x)$ and $p_{i',i}(x)$ will be elements of $E_{i'}$ mapping to the same element of E, hence (again by the fact that E is the direct limit of our given system), if we take an i'' sufficiently high above i', we will have

$$p_{i'',i'} \theta_i h_i(x) \;=\; p_{i'',i'} p_{i',i}(x) \;=\; p_{i'',i}(x).$$

Moreover, if we choose i'' so that the above relation holds for all members of a finite generating set for P_i, it will hold for all $x \in P_i$. Renaming things, we conclude that for each $i \in I$ we can choose an $\alpha(i) \geq i$ (what we called i'' above), and a homomorphism

$$\varphi_i : F_i \;\to\; P_{\alpha(i)}$$

(called $p_{i'',i'} \theta_i$ above) such that

$$\varphi_i h_i \;=\; p_{\alpha(i),i}.$$

Let us now define a relation $\leq'$ on I by

$$j \geq' i \text{ if and only if } j = i \text{ or } j \geq \alpha(i).$$

It is easy to verify that this is a partial ordering, and again makes I a directed partially ordered set. Further, for $j \geq' i$ let us define $f_{j,i} : F_i \to F_j$ by

$$f_{j,i} \;=\; \begin{cases} \mathrm{id}_{F_i} & \text{if } j = i, \\ h_j \, p_{j,\alpha(i)} \, \varphi_i & \text{if } j \geq \alpha(i). \end{cases}$$

You should now be able to verify that these maps make the free modules F_i a directed system, with E as its direct limit. This completes the proof of

Theorem XVI.3.c2 (D. Lazard; cf. Lang, Exercise XVI:L13)**.** *Let E be a right module over a ring R. Then the following conditions are equivalent.*

(a) *E is flat.*

(b) *Every homomorphism from a finitely presented right R-module P into E factors through a free module.* (*Equivalently, in the ad hoc language introduced earlier in this discussion, every finite family of*

relations holding among a finite family of elements of E *holds ''for good reason''*).

(c) E *is the direct limit of a directed system of free right modules of finite rank.* □

The above proof of (b)⇒(c) follows the Bourbaki reference given in Lang. The proof in Lazard's original paper (*Sur les modules plats*, Comptes Rendus Acad. Sci. Paris, **258** (1964) pp.6313-6316) is sketched as Exercise XVI.3:3.

Let us note that from one of the relatively easy parts of the above lemma, the implication (a)⇒(b) (equivalent to Lemma XVI.3.10), we get the

Corollary XVI.3.c3 (D. Lazard)**.** *If a flat module* E *is finitely presented, it is projective.*

Proof. The identity map of E is in this situation a homomorphism from a finitely presented module to a flat module, hence it factors through a free module F:

$$E \xrightarrow{b} F \xrightarrow{a} E \quad \text{with} \quad ab = \mathrm{id}_E.$$

It follows that the other composite, ba, is a retraction of F onto a direct summand isomorphic to E. □

Incidentally, suppose that E is a finitely generated projective module, say the image of an idempotent endomorphism e of a free right module ${}^nR$. By Lang's Proposition XIV.3.1(iii), we know that E is flat, and we may ask how to get an explicit directed system of free modules of finite rank having E as direct limit. Amusingly, such a system is given by

$${}^nR \xrightarrow{e} {}^nR \xrightarrow{e} {}^nR \xrightarrow{e} \ldots .$$

Now back to the initial pages of this section of Lang.

P.613, Proposition XVI.3.2 [=]: In this proposition, we must assume R commutative. (As noted in my comments on the end of §II.4, there are various noncommutative concepts of ''localization''. Some of these do and others don't have the property that a localization of a ring R is flat as a right R-module; Exercise XVI.3:11 gives examples.) For the definitions of the symbols used in (ii) and (iii), see Exercises III:L9-10 (pp.167-168).

The proofs of these three statements don't seem to me as obvious (in terms of the material given so far) as Lang claims. I would guess that the proof of statement (i) that Lang had in mind was to combine Exercise III:L9(b) with the fact that $S^{-1}M \cong M \otimes_R S^{-1}R$. However, that isomorphism hasn't been noted yet; in fact, the right context for it is the next section.

A proof appropriate here would be the following. Let $u\colon X \to S$ be a homomorphism from a free abelian monoid X (written multiplicatively) onto S. Now a free abelian monoid can be ordered by *divisibility*, and is directed under this ordering. For every pair of elements related under this ordering, $x \leq xy \in X$, let us define $f_{xy,x}\colon R \to R$ to be the R-module homomorphism given by multiplication by $u(y)$. This defines a directed system of R-modules (all equal to R) indexed by X, and it is easy to verify that the direct limit of this system can be identified with $S^{-1}R$ by sending the copy of $r \in R$ in the member of this system indexed by $x \in X$ to $u(x)^{-1}r$. Hence $S^{-1}R$, being a direct limit of free modules, is flat.

In statement (ii), the implication that if M is flat then so is $M_{\mathbf{p}}$ is true in the more general form ''if M is flat, then so is $S^{-1}M$ for any multiplicative subset S''. (Recall that $M_{\mathbf{p}}$ means $(R-\mathbf{p})^{-1}M$.) This can be proved exactly like (i), by writing $S^{-1}M$ as a direct limit of copies of M. For the other direction, first note that localization commutes with tensor products, i.e.,

$$(S^{-1}M) \otimes_R E \;\cong\; S^{-1}(M \otimes_R E) \;\cong\; M \otimes_R (S^{-1}E). \tag{c53}$$

This can easily be seen from the above direct limit description of localization. Now if M is not flat, find a short exact sequence whose exactness is not preserved under $M \otimes_R -$. Then note that by

Exercise III:L10(b), there must be some prime $\mathbf{p}$ such that this nonexactness remains true on localizing at $\mathbf{p}$; and finally, use (c53) to turn this nonexactness into the statement that $M_{\mathbf{p}}$ is not flat.

Turning to (iii), to get the ''if'' direction, assume F is torsion free, regard it as the directed union of its finitely generated submodules, and note that this expresses it as a direct limit of finitely generated torsion free modules, which by the theory of finitely generated modules over a PID are free. For the ''only if'' direction, observe that if F has torsion, then the calculation by which we showed that the $\mathbf{Z}$-module $\mathbf{Z}/p\mathbf{Z}$ was non-flat can be adapted to show M non-flat. Namely, if F has a nonzero element annihilated by a nonzero $r \in R$, then tensoring with M does not preserve the one-one-ness of the map $R \xrightarrow{r} R$. This argument, indeed, gives the more general statement on the last two lines of the page. (Alternatively, we can get this using Lazard's Theorem, i.e., Theorem XVI.3.c2 above, together with the observation that a free module over an integral domain is torsion-free, hence so is a direct limit of such modules.)

P.614, first two paragraphs [<].

Examples of non-flat torsion-free modules.

Lang describes several classes of examples, but these assume a nontrivial amount of number theory and/or commutative algebra; so let me give some cases where non-flatness can be verified ''by hand''.

First, let R be a polynomial ring $k[x,y]$ in two indeterminates over a field. (This is not the $k[x,y]$ of Lang's first paragraph, since his x and y satisfy a polynomial relation.) Now it follows from unique factorization in R that in a free R-module, if we have two elements ξ and η such that $\xi y - \eta x = 0$, there must be an element α such that $\xi = \alpha x$ and $\eta = \alpha y$; hence the same will be true in any direct limit of free modules, i.e., any flat module. Within R itself, if we take $\xi = x$, $\eta = y$, then of course such an element is given by 1. But in the ideal (x,y) of R, there is no such element, hence that ideal is not flat as an R-module. (In other words, x and y satisfy the indicated module-relation ''for good reason'' in R, but not in (x,y).)

This argument can be formulated so as not to use the hard part of Lazard's Theorem, but only the implication (a)$\Rightarrow$(b). For this shows that if (x,y) were flat, we could factor its identity map through a free module. Comparing the coordinates of the images of x and y in that free module, we can deduce by unique factorization that those images have the forms αx and αy respectively; then, considering the image of α under the map from this free module back into (x,y), we get a contradiction.

Can we give an explicit example where $(x,y) \otimes_R -$ does not preserve exactness? Yes; tensor (x,y) with the exact sequence

$$0 \to (x,y) \to R \to R/(x,y) \to 0.$$

The element $x \otimes y - y \otimes x$ of $(x,y) \otimes_R (x,y)$ belongs to the kernel of the map into $(x,y) \otimes_R R \cong (x,y)$, but is nonzero. The former fact is immediate; to show the latter, make k an R-module by letting x and y act trivially, and define an R-bilinear map from $(x,y) \times (x,y)$ to the direct sum of four copies of this module, so as to send each member of $\{x,y\} \times \{x,y\}$ to a different basis element. One finds that the induced R-linear map from $(x,y) \otimes_R (x,y)$ into that k-vector space takes $x \otimes y - y \otimes x$ to a difference of basis elements. (One copy of k would have sufficed instead of four, but this would have required using a contrived-looking bilinear map, while the map used in this case looked, I hope, natural.) The same method works for the power series ring $k[\![x,y]\!]$. (Lang gives this example in the second paragraph of this page, with a different argument.)

From the point of view of algebraic geometry, the above examples used rings whose spectra have dimension > 1. (The student who is not familiar with algebraic geometry should not worry; this is just an aside to those who are. The ring-theoretic arguments below will be self-contained. My later reference to the concept of integral closure can likewise be ignored by the reader not familiar with that concept.) There are also rings whose spectra are 1-dimensional, but have a singularity at some point that make them ''fat'' enough there to get a similar example. For example, consider the subring $k[x^2, x^3] \subseteq k[x]$. One can

show that its ideal (x^2, x^3) is nonflat as a module by using x^2 and x^3 in place of x and y in the arguments for the preceding example. (Instead of unique factorization, which does not hold, one must consider explicitly the factorization properties of powers of x in this ring; but this is not difficult; the equation $x^2 \cdot x^3 = x^3 \cdot x^2$, when looked at as a module relation, still holds without ''good reason'' in the ideal.) Note further that if we regard $k[x]$ as a module over its subring $k[x^2, x^3]$, then the linear map $f(x) \mapsto x^2 f(x)$ is an isomorphism between this module and the above ideal of $k[x^2, x^3]$; so having proved that ideal non-flat, we can say that the ring $k[x]$, which is the integral closure of $k[x^2, x^3]$, is non-flat as a module over that subring. Thus, this is an example in the spirit of the one with which Lang begins this paragraph.

The above subring of $k[x]$ can be described as consisting of those functions whose derivative at $x = 0$ is zero; from the point of view of algebraic geometry, its spectrum is the line ''pinched'' at that point. Another example of non-integral-closure is gotten by taking the line and ''tying together'' the two points $x = 1$ and $x = 0$, getting the ring $\{ f \in k[x] \mid f(0) = f(1) \}$. A non-flat ideal of this ring is given by the set of polynomials whose common value at those two points is 0; in showing non-flatness, one may use the elements $x^2(x-1)$ and $x(x-1)^2$ in the role of ''x and y''. Again, $k[x]$ is R-module-isomorphic to this ideal, this time by $f(x) \mapsto x(x-1) f(x)$.

Rather than going through separate but parallel arguments for nonflatness in each of the above cases as sketched above, one can handle them all with one lemma; this is done in Exercise XVI.3:4. Exercise XVI.3:9 gives another condition which quickly shows the non-flatness of the last two examples.

P.614, Lemma XVI.3.3 [>]: Note that this result is easy when F is projective, since the given exact sequence splits; this illustrates the fact that flat modules can be considered ''close to'' projective. An alternative proof of the lemma, which uses Lazard's Theorem to reduce to the case where F is in fact free, and hence projective, is sketched in Exercise XVI.3:5 (which also does the same for Proposition XVI.3.4 on the next page).

In Lang's proof, note that he does not show the full diagram of the snake lemma, but only those parts that are relevant to the argument at hand. He will follow this practice in subsequent proofs as well.

P.615, Proposition XVI.3.4 [>]: Strictly speaking, for the phrase ''*More generally*'' (preceding the last display) to be correct, the last line should say that if $F^2, \dots, F^n$ are flat, then F^0 is flat if and only if F^1 is. This is true, and may be proved by induction from the first part of the proposition: Let $n > 2$, let us assume inductively that the above statement is true for smaller n, and let K denote the kernel of the map $F^2 \to F^3$, equivalently, the cokernel of the map $F^0 \to F^1$, in the given exact sequence. Thus, that sequence decomposes into two exact sequences

$$0 \to F^0 \to F^1 \to K \to 0, \qquad 0 \to K \to F^2 \to \dots \to F^{n-1} \to F^n.$$

Our inductive hypothesis, applied to the second sequence, shows that K is flat, and the first part of the proposition now shows the asserted conclusion about F^0 and F^1.

The above is what Lang means by the second paragraph on the next page. The longer paragraph which precedes it proves the first assertion of the result. In following that proof, note that where, just below the diagram, Lang says ''The 0 on the right'' he means the 0 at the top; and that for the remainder of the proof, he is using what the snake lemma says about the kernels of the three vertical maps. The importance of the 0 on the left in the first row is that it shows that the map of kernels above is also one-to-one.

P.617, five lines above bottom [=]: The phrase ''it becomes 0 already in a finite subsum'' is vague. A more precise formulation of the argument is that any element of the indicated kernel is a finite sum $z = \Sigma\, x_j \otimes y_j$ $(x_j \in F,\ y_j \in E')$, and each $y_j \in E' \subseteq E$ is a sum of elements from finitely many of the E_i. If we write $E_{I_0} \subseteq E$ for the sum of the finitely many E_i that are thus involved in expressing z, and let $E'_{I_0} = E' \cap E_{I_0}$, then we can show that the element $\Sigma\, x_j \otimes y_j \in F \otimes E'_{I_0}$ is in the kernel of the map to $F \otimes E_{I_0}$, and is nonzero if z is, because one can map it to z. Thus, if the result is true for finite direct

sums E_{I_0}, it is true for arbitrary direct sums E.

But an easier approach is to forget about dealing with individual elements, and note that the exact sequence $0 \to E' \to E$ is the direct limit, over the directed partially ordered set of all finite subsets $I_0 \subseteq I$, of the exact sequences $0 \to (E' \cap E_{I_0}) \to E_{I_0}$, and that direct limits respect tensor products and exactness. In either case, we are reduced to finite direct sums, for which Lang's proof is fine, except for a pair of misprints, noted below:

P.618, end of proof of Lemma XVI.3.6 [>]: Now observe that if an R-module F is ''R-flat'', then by the above lemma, it is flat with respect to any *free* R-module, hence by the preceding lemma, it is flat with respect to any homomorphic image of a free R-module, i.e., with respect to any R-module; i.e., it is flat. As the converse is trivially true, we see that flat is equivalent to R-flat. This is the culmination of the preceding study of ''E-flatness'', and is the content of the second assertion of the next result, Proposition XVI.3.7. (For the noncommutative version, we should assume F a right R-module, and let $\mathbf{a}$ range over all left ideals of R.)

Now the exactness condition associated with the short exact sequence of that assertion is equivalent to saying that the map $F \otimes \mathbf{a} \to F$ is one-to-one. The image of this map is $F\mathbf{a}$, so this says that the natural map $F \otimes \mathbf{a} \to F\mathbf{a}$ is an isomorphism, which is the first assertion of the proposition (adjusted to make modules we are testing for flatness right modules).

The above is essentially Lang's proof of the proposition, but I think more clearly stated.

P.619, Lang's **Remark on abstract nonsense [**<**]**: Skip this (except the last one-sentence paragraph!) unless you know about, and like, abelian categories.

If you do read it, note that there is an unnamed third abelian category, which we could call $\mathcal{C}$, in which T is assumed to take values; and that ''right exactness'' is the name of the property proved for $F \otimes -$ in Proposition XVI.2.6 (p.610).

P.619, Theorem XVI.3.8 [=]: Lang will first prove this in the case where R is Noetherian, then show how to modify the proof to work under the weaker assumption that M is finitely presented, and finally, he will give a different proof for the general case.

Concerning his proofs in the first two cases, I will merely mention that Lemma XVI.3.9 can be considered an immediate consequence of an easily proved but powerful result given in a later chapter, Schanuel's Lemma (Lemma XXI.2.4).

The big tool that he brings to bear to get the general case of Theorem XVI.3.8 is Lemma XVI.3.10, which we have already discussed. His explanation of how to get the theorem from that lemma (next-to-last paragraph of p.621) is sketchy, so I will give a proof here.

Suppose M, R and $x_1, \ldots, x_n$ are as in the theorem, where we will as usual understand M to be a right module. By Nakayama's Lemma (say the version given as Lemma X.4.3), $x_1, \ldots, x_n$ generate M. Let $f\colon F \to M$ be a homomorphism from a free module of rank n onto M taking a basis of F to $x_1, \ldots, x_n$. The fact that $x_1, \ldots, x_n$ are linearly independent modulo $M\mathbf{m}$ tells us that $\mathrm{Ker}(f) \subseteq F\mathbf{m}$. We wish to show that this kernel is 0.

So say $y \in \mathrm{Ker}(f)$. Then f factors

$$F \to F/yR \to M$$

Here the right-hand map is from a finitely presented module to a flat module, so by our translation of Lemma XVI.3.10, it factors through a free module G. This means that our map f factors through G

$$F \xrightarrow{g} G \xrightarrow{h} M$$

in such a way that $y \in \mathrm{Ker}(g)$. On the other hand, since f is surjective, we can factor h through it (since the domain of h is free):

$$F \xrightarrow{g} G \xrightarrow{e} F \xrightarrow{f} M.$$

Thus $feg = f$, so the endomorphism $eg - \mathrm{id}_F$ of F has range in $\mathrm{Ker}(f) \subseteq F\mathbf{m}$, so the image under

eg of a basis $\mathbb{B}$ of F is congruent elementwise to $\mathbb{B}$ modulo $F\mathbf{m}$, so by Theorem X.4.4, $eg(\mathbb{B})$ is a basis. In particular, its elements are linearly independent, so eg is one-to-one. But $y \in \mathrm{Ker}(g)$, so $y = 0$, as required.

P.622, [<]: Skip this unless you have read Chapter XX or have the equivalent background.

Re §XVI.4. Extension of the base.

P.623, beginning of §XVI.4 [<]: What Lang does here is valid in the noncommutative context (say for left R-modules, since he is tensoring on the left with extension rings), up to a point on the next page where I will note the contrary.

Before we go further, let us play some

Games with adjoints

Given a ring homomorphism $f: R \to R'$, note that there is a very simple construction taking left R'-modules to left R-modules: Regard the underlying abelian group of the given R'-module F as an R-module by defining rx (for $r \in R$, $x \in F$) to be $f(r)x$. This is called *restriction of scalars* from R' to R (a term adapted from the special case where f is the inclusion of a subring R in a larger ring R'). You should not find it hard to verify that the construction of *extension of scalars*, described by Lang, is left adjoint to restriction of scalars. You should write out the universal property that this implies.

The next two results that Lang will prove can both be looked at as immediate consequences of the fact that composites of adjoint functors are adjoints of the composite functors (Lemma II.3.c3). Indeed, given a ring homomorphism $f: R \to R'$, consider the functors

$$R'\textbf{-modules} \xrightarrow{\text{restriction of scalars}} R\textbf{-modules} \xrightarrow{\text{forget module structure}} \textbf{Sets},$$

whose composite is the forgetful functor from R'-modules to sets. The left adjoints of the above two functors are

$$R'\textbf{-modules} \xleftarrow{\text{extension of scalars}} R\textbf{-modules} \xleftarrow{\text{free } R\text{-module}} \textbf{Sets},$$

so these compose to the left adjoint to the above composite, which is the free R'-module functor. Thus, extension of scalars takes free modules to free modules, which is Proposition XVI.4.1.

Likewise, if we have two ring homomorphisms $R \to R' \to R''$, and we consider successive restriction of scalars from R''-modules to R'-modules to R-modules, then the above observation on composites of adjoint functors gives the result that Lang on the next page calls transitivity of extension of base.

Going back to the definition of extension of scalars, let us note that where Lang has described ''by hand'' how to make $R' \otimes_R E$ into an R'-module, there is a neat formal approach. Given a homomorphism $R \to R'$, let us regard R' as an (R', R)-bimodule, using its natural left module structure over itself, and getting the right R-module structure by starting with its right module structure over itself, and then restricting scalars. Then extension of scalars can be regarded as the operation of tensoring left R-modules with this (R', R)-bimodule, which we know gives left R'-modules.

Suppose instead that we look at R' as an (R, R')-bimodule. Then we can tensor a left R'-module E with it to get a left R-module. What is this module? Tensoring an R'-module with R' is essentially the identity functor; so this construction associates to any left R'-module E the ''same'' object, but considered as a left R-module. It is not hard to check that this gives the restriction of scalars functor!

We saw above that restriction of scalars was right adjoint to extension of scalars; but here we find that it can be expressed as a tensor product, hence as a left adjoint functor. What is its right adjoint? You can verify that it is the functor associating to every left R-module E the abelian group $\mathrm{Hom}_R(R', E)$, where by this we understand the set of homomorphisms as left R-modules, and where we make this set a left R'-module with the help of the *right* R'-module structure of R'. (I.e., if $f \in \mathrm{Hom}_R(R', E)$, and we write f, according to our convention, to the right of its argument, then for any $s \in R'$, we define sf by $x(sf) = (xs)f$.)

Thus, the restriction-of-scalars functor belongs to the rather special class of functors that have both left

and right adjoints. (Another such functor is the forgetful functor from monoids to groups.)

P.624, a third of the way down, where Lang writes ''If E has a multiplicative structure'' [<]: From this point to the bottom of the page, R should be assumed commutative.

P.625, Proposition XVI.4.2 [=]: The two results that Lang labels ''Base change'' and ''Transitivity'' might better be labeled ''Extension of scalars'' and ''Restriction of scalars''.

Both of these results are easy to get using associativity of the tensor product. Unfortunately, Lang has only noted this associativity for modules over a common commutative ring, and this is not quite enough. The useful result here is associativity of tensor products of bimodules: If E is a right R-module, F an (R, S)-bimodule, and G a left S-module, then

$$(E \otimes_R F) \otimes_S G \;\cong\; E \otimes_R F \otimes_S G \;\cong\; E \otimes_R (F \otimes_S G),$$

and, of course, if either or both of the modules E, G is in fact a bimodule, these isomorphisms respect the resulting module or bimodule structures on these tensor products. If we now use the characterizations of extension and restriction of scalars as tensoring operations, then we immediately get noncommutative versions of the above two results, namely that if $R \to S$ is any ring homomorphism, then extension of scalars takes flat left R-modules to flat left S-modules, while if S is flat as a left R-module, restriction of scalars takes flat left S-modules to flat left R-modules. In each case, to verify that the tensor product modules in question are flat, ''test'' for flatness by tensoring them with a one-one-map of the appropriate sort of module, and use associativity of the tensor product to deduce a positive answer to this test. (Why is a hypothesis on S needed in the case of restriction, but no corresponding hypothesis needed in the case of extension?)

Re §XVI.5. Some functorial isomorphisms.

P.625, lines immediately below the diagram [=]: What Lang calls an ''additive category'' is more commonly called (and I will call it) an **Ab**-category. Here **Ab** is the standard symbol for the category of abelian groups; an **Ab**-category means a category whose hom-sets are objects of **Ab**, and such that for every morphism f, the composition operations $f \circ -$ and $- \circ f$ act by morphisms of **Ab**. The term ''additive category'' is commonly used for **Ab**-categories in which every pair of objects has a coproduct, i.e., a direct sum. Most of the **Ab**-categories that come up naturally satisfy a still larger set of important properties, which are abstracted in the definition of *abelian category*, Lang, p.133. However, the simpler concept of **Ab**-category is all that he needs for the definitions he makes here.

I don't know whether the term ''additive functor'' is standard; I will follow Lang in this usage.

To see that not every naturally occurring functor among **Ab**-categories is additive, let R be a commutative ring, and consider the functor from R-modules to R-modules taking E to $E \otimes_R E$.

P.626, Proposition XVI.5.1 [~]: This result and its proof are valid for right or left modules over not necessarily commutative rings. More generally, the argument shows that if a morphism of additive functors between module categories (not necessarily modules over the same ring) is an isomorphism on some family of modules, it is an isomorphism on their direct sum. Still more generally, one could prove such a result for additive functors among **Ab**-categories, but one would first need to develop tools for studying products and coproducts in such categories, rather than working with elements as Lang does here.

P.627, Corollary XVI.5.3 [~]: Here let's keep Lang's assumption that R is commutative. (If we want a noncommutative version of this result, we need a context where it first makes sense for the free module of rank 1; but I don't see any very natural contexts where the sets $L(R, R)$ have module structures that allow us to take tensor products, and such that the two sides come out isomorphic.)

P.628, Corollary XVI.5.4 [~]: Calling this a ''functorial isomorphism'' is a little sticky, because $L(E', E)$ is covariant in the first variable and contravariant in the second variable, hence one cannot make $L(E, E)$ into either a covariant or a contravariant functor in its one variable.

A way to get around this is to restrict attention to the category whose objects are our modules, but

whose morphisms are the *isomorphisms* among these modules. In that case, by taking isomorphisms to their inverses, one can interchange covariant and contravariant functors, and indeed, given an isomorphism $f: E \cong E'$, one gets an isomorphism $\mathrm{End}(E) \cong \mathrm{End}(E')$, namely $g \mapsto fgf^{-1}$. As a construction on this restricted category, the isomorphism of this corollary is in fact functorial.

P.628, Corollary XVI.5.5 [~]: This is valid for right modules over an arbitrary ring R. Recalling that $E^{\vee}$ should then be considered a left R-module, we see that both sides of the isomorphism can be regarded as abelian groups.

P.629, both results [<]: Assume R commutative here.

Re §XVI.6. Tensor product of algebras.

Assume R commutative throughout this section, but note that the R-algebras A that Lang is considering need not be commutative.

P.630, Proposition XVI.6.1 [>]: This result, that in the category of commutative R-algebras the tensor product gives the coproduct, is actually a case of a characterization of the tensor product of two not-necessarily-commutative R-algebras A and B: namely, that $A \otimes_R B$ is universal among R-algebras given with homomorphisms of A and B into them such that every element of the image of A commutes with every element of the image of B.

P.631, line beginning **Graded Algebras.** [=]: There is no need to assume G commutative. Probably Lang's reason was that the most frequent grading monoids are the nonnegative integers, the integers, and $\mathbf{Z}/2\mathbf{Z}$, all of which are written additively, so he wants to use additive notation. But gradings by noncommutative monoids, especially noncommutative groups, are also important, and people who work with graded rings are comfortable with both additive and multiplicative notation.

P.631, last paragraph [~]: Actually, if A is $\mathbf{Z}$-graded, then the graded algebra Lang here denotes A_{su} is isomorphic to A; but when A is $\mathbf{Z}/2\mathbf{Z}$-graded, they can be nonisomorphic (Exercise XVI.6:1). On the other hand, the ''super tensor product'' discussed on the top half of the next page is in general not isomorphic to the ordinary tensor product for either grading group.

P.632, end of §XVI.6 [>]. You might like to try your hand at

Open Question XVI.6.c1. *If k is an algebraically closed field, and D, E are division algebras over k, must $D \otimes_k E$ be a k-algebra without zero-divisors?*

The answer is ''yes'' if D and E are commutative (or even if one of them is), ''no'' if we drop the assumption that k is algebraically closed, also ''no'' if we keep that assumption but weaken the hypothesis on D and E from ''division algebras'' to ''algebras without zero-divisors''. (Cf. my paper *Zero-divisors in tensor products*, pp.32-82 of Springer Lecture Notes in Mathematics, v.545, R. Gordon, ed., *Noncommutative ring theory, papers presented at the international conference at Kent State University, 4-5 April 1975,* Springer Verlag, 1976.)

Re §XVI.7. The tensor algebra of a module.

I generally teach the subject of tensor algebras (this section), symmetric algebras (next section) and exterior algebras (§XIX.1) from notes of mine, which can be found through my web page. I will give here brief notes on Lang's presentation of this material.

P.633, line 10 [=]: I'm not sure what Lang means by ''the map of §1''. Perhaps ''the $r=1$ case of $T(f,\dots,f)$''.

P.633, italicized sentence in middle of page [=]: When he calls $T(f)$ a ''homomorphism of degree zero'', Lang is using terminology which he has not introduced: If $M = \oplus_G M_r$ and $N = \oplus_G N_r$ are two modules, written as direct sums indexed by an abelian group G, then for $d \in G$ a ''homomorphism $M \to N$ of degree d'' is defined to mean a module homomorphism $M \to N$ which for each r carries M_r into M_{r+d}.

(One can set up similar language for G noncommutative, but one must then distinguish between

homomorphisms carrying M_r into M_{rd} and homomorphisms carrying M_r into M_{dr}.)

P.633, next paragraph [=]: There's no reason for assuming E to have finite rank; the objects which are sometimes called, as Lang does, noncommutative polynomial rings, and more often called free associative algebras, are generally studied without such an assumption.

The free associative algebra on a set X of indeterminates is most often denoted $R\langle X\rangle$, and the tensor algebra on a module E is often denoted $R\langle E\rangle$. The former is universal among R-algebras given with set-maps of X into them; the latter among R-algebras given with R-module homomorphisms of E into them. In other words, the two functors denoted $R\langle -\rangle$ are left adjoint to the forgetful functor from R-algebras to sets, and to the forgetful functor from R-algebras to R-modules, respectively.

P.634, third line, through end of the section [~]: This is devoted to proving Proposition XVI.7.2. I'm not sure why Lang considers this result important; perhaps he will make use of it in some section I haven't read. I will generally not assign this page.

P.634, end of §XVI.7 [>]: A large amount of interesting ring theory has been done involving free associative algebras. See P. M. Cohn, *Free Rings and Their Relations*, Academic Press, 1985. My doctoral thesis (though it proved most results in more general contexts) was inspired by questions about these objects. They are also the starting-point in the study of structures of algebras presented by generators and relations, on which I usually spend a few lectures, assigning reading from the first few sections of my paper, *The diamond lemma for ring theory*, Advances in Mathematics **29** (1978) 178-218.

There is a generalization of the concept of tensor algebra. If R is a not-necessarily-commutative ring, one defines an *R-ring* to mean a ring given with a homomorphism of R into it; thus, an R-ring has a natural structure of (R,R)-bimodule. Now if E is any (R,R)-bimodule, one can define $R\langle E\rangle$ to be the direct sum over r of the r-fold tensor products $E\otimes ... \otimes E$ of bimodules, where again the 0-fold product is defined to be R. This is universal among R-rings with (R,R)-bimodule homomorphisms of E into them.

Re §XVI.8. Symmetric products.

P.635, bottom [>]: $S(E)$ is universal among *commutative* R-algebras given with R-module homomorphisms of E into them.

P.636, bottom, Proposition XVI.8.2 [~]: This is true without the assumption that E' and E'' be free of finite rank. (Though freeness is needed for the second proof that Lang suggests, based on identifying $S(E)$ with a polynomial algebra.)

P.647, statement of Lemma XVII.3.1 [=]: ''*Let* $x\in R$'' should be ''*Let* $x\in E$''.

P.689, third line after boxed display [>]: Lang hasn't introduced the concept of adjoint functor, so this remark can be ignored if you haven't seen it defined. However, I discussed some cases of the concept above in the notes on Lang's p.607 and p.623. For a full motivated development, with many varied examples, see §§7.1-7.3 of my Math 245 notes.

P.735, sentence before final Example [=]: It is not clear to me what Lang means by this, since the values of this map, 1-dimensional vector spaces, don't form an abelian group. Perhaps he means this to be an intuitive idea rather than a literal assertion.

Appendix 2. Some Set Theory.

Re §A2.1. Denumerable sets.

P.876, last line of Proof of Proposition A2.1.1 [=]: To prove that the map $n \mapsto k_n$ enumerates D, one shows by induction that for every n, $\{k_1, \dots, k_n\}$ includes all members of D that are $\leq n$.

P.877, proof of Proposition A2.1.3 [>]: This proof uses the Axiom of Choice, which Lang considers so obvious it need not be mentioned, but which I discuss in the handout ''The Axiom of Choice, Zorn's Lemma, and all that'' to be read in connection with §A2.2. One could prove this proposition in a way that hides the fact that it requires this Axiom, using a sequence of single choices (each of which could be done without the Axiom), and an argument ending with ''and so forth'' (which implicitly means one is making infinitely many choices, which requires the Axiom). Lang, however, makes all the choices in one step, which is just how a well-stated proof using the Axiom of Choice should go. The reader not familiar with the Axiom of Choice need not worry about this now, since this is not a course in set theory, but should simply be alerted that a set-theoretic assumption we will discuss later has been used implicitly here.

At the point of this proof where Lang says ''By induction'', it would be more proper to say ''by recursion'': an inductive proof proves the n+1st of a sequence of *statements* from the nth; a recursive construction constructs the n+1st of a sequence of *objects* from the nth.

Re §A2.2. Zorn's Lemma.

P.879, lines 3 and 4 [=]: ''If the ordering satisfies this additional condition ...''. Delete this sentence. (What is said unclearly here is said again clearly near the bottom of the same page.)

P.879, 2/3 down, definition of **induced** ordering [>]: Note that Examples 1 and 2 can be obtained from Example 3 by taking such induced orderings.

P.880, middle, to end of section on p.884 [<]: I recommend that you read my handout *The Axiom of Choice, Zorn's Lemma, and all that* for this material, then simply look over this section of Lang to familiarize yourselves with his viewpoint. The module-theoretic result stated on p.880, and proved there using Zorn's Lemma, is discussed in the notes on §I.7 in this Companion, so if you are reading this material before that section, don't worry about it yet.

Re §A2.3. Cardinal numbers.

P.885, first half of page [<]: This is a bit disorganized and repetitive, but wade through it and absorb the elementary facts noted.

P.886, line 9, ''Note first ...'' [=]: This is by Proposition A2.1.3 (p.877).

P.887, Corollary A2.3.4 [=]: The hypothesis ''*Let A be an infinite set*'' from the preceding theorem is still assumed.

P.891, end of §A2.3 [>]: Lang has introduced the symbol ''card(X)'', and given meaning to equalities and inequalities of these symbols, but has not defined any mathematical entities that the symbols represent. Set-theorists define certain sets called *cardinals*, which are canonically chosen ''sample sets'' of every cardinality, and given a set X, they define $\mathrm{card}(X)$ to be the unique cardinal with which X can be put into bijective correspondence. I often sketch the construction of the cardinals in class; it depends on another construction, that of the *ordinals*, which are representatives of the isomorphism classes of well-ordered sets. Though there is exactly one *finite* ordinal of each finite cardinality, in the infinite case, the correspondence becomes many-to-one, e.g., the *denumerable* ordinals include ω, ω+1, ... , $\omega+\omega$, ... The cardinals are defined as those ordinals that are *least* among the ordinals of their cardinality. For a more detailed sketch, see Chapter 4 of my Math 245 notes in particular, §4.5. For a full development, see a text on set theory.

P.891, Theorem A2.3.10 and Corollary A2.3.11, [>].

Some inequalities for cardinals (optional for my 250A).

Consider the following inequalities which are true for nonnegative integers:

(c54) $a_i < b_i \quad \Rightarrow \quad \Sigma\, a_i < \Sigma\, b_i \qquad (a_i \geq 0;\ \ 0 \leq i < n;\ \ n \geq 1)$

(c55) $a_i < b_i \quad \Rightarrow \quad \Pi\, a_i < \Pi\, b_i \qquad (a_i \geq 1;\ \ 0 \leq i < n;\ \ n \geq 1),$

(c56) $a_i \leq b_i \quad \Rightarrow \quad \Sigma\, a_i < \Pi\, b_i \qquad (a_i \geq 3;\ \ 0 \leq i < n;\ \ n \geq 2).$

As noted in A2:L8, there are also natural arithmetic operations on cardinals – but one finds that none of (c54)-(c56) hold for possibly infinite sums and products of possibly infinite cardinals! A counterexample to (c54) can be seen from the fact that any infinite sum of 1's is equal to the corresponding infinite sum of 2's (or of $\aleph_0$'s), although $1 < 2$. A counterexample to (c55) is similarly gotten by comparing countable products of 2's and 4's. To see that (c56) fails, take all a_i's and b_i's to be a fixed infinite cardinal, and take the sum and product over a finite set. Having resigned outselves to these features of cardinal arithmetic, it comes as a surprise that the law

(c57) $\alpha_i < \beta_i \quad \Rightarrow \quad \Sigma\, \alpha_i < \Pi\, \beta_i$

holds for arbitrary (zero, positive-integer, or infinite) cardinals α_i and β_i, indexed by an arbitrary (empty, finite nonempty, or infinite) set I. The theorem and corollary that you have just read in Lang, saying that for any set I one has $\mathrm{card}(I) < \mathrm{card}(2^I)$, is the case of this where all $\alpha_i = 1$ and all $\beta_i = 2$. The inequality (c57), known as the König-Zermelo Theorem, is in fact proved in essentially the same way as this special case, if the proof of that case is appropriately formulated. Namely, in proving the special case, consider a function f from the disjoint union U of an I-tuple of 1-element sets to the I-fold product P of the set $\{0, 1\}$. For each $i \in I$, there is just *one* element u_i in the ith member of the union U, but there are *two* elements of $\{0, 1\}$ which are the possible values of the ith term of a member of P, hence, since $1 < 2$, for each i we can choose a member of $\{0, 1\}$ which is not the ith term of $f(u_i)$. From this one deduces that f is not surjective. You should easily be able to adapt that argument to prove (c57).

As another application of the König-Zermelo Theorem, consider any increasing sequence of cardinals $\alpha_0 < \alpha_1 < \ldots$ indexed by the positive integers, and let $\alpha = \Sigma\, \alpha_i$. Then

$$\alpha \;=\; \Sigma\, \alpha_i \;<\; \Pi\, \alpha_{i+1} \;\leq\; \Pi\, \alpha \;=\; \alpha^{\aleph_0}.$$

In the language of Exercise A2:L13 (p.893), this says $\alpha < \alpha'$. Since α can be taken greater than any cardinal β (by letting $\alpha_0 = \beta$ and $\alpha_{i+1} = 2^{\alpha_i}$ for $i \geq 0$), this answers the question raised in brackets in that exercise. (Solovay showed me the König-Zermelo Theorem, and I was the ''grapevine'' that communicated it to Lang.)

Incidentally, the proof of the result $\mathrm{card}(I) < \mathrm{card}(2^I)$ is itself a generalization of the observation that a set theory which allows one to speak of ''the set of all sets'' leads to a contradiction. Namely, the contradiction when we have a ''set of all sets'' arises by noting paradoxical properties of its subset, ''the set of all sets that are not members of themselves''; in the proof that $\mathrm{card}(I) < \mathrm{card}(2^I)$ one considers a map $f\colon I \to 2^I$, and effectively notes that the set of all $i \in I$ that are not members of $f(i)$ cannot have the form $f(j)$ for $j \in I$; hence f is not surjective.

The paradox about the set of all sets that are not members of themselves can in turn be considered a descendent of paradoxes about the truthfulness of people who assert that they always lie, etc., which go back thousands of years!

Re §A2.4. Well-ordering. The material in this brief section is also included in the ''Axiom of Choice, etc.'' handout.

P.892, line 2 [=]: Both occurrences of A in this line should be S. (The A on the preceding line is correct.)

P.892, Theorem A2.4.1 [>]: The word ''nonempty'' in the result is unnecessary – the empty set can be well-ordered, too. In the fifth line of the proof, an ''initial segment'' of a totally ordered set X means a subset Y such that whenever $y \in Y$ and $z < y$ we also have $z \in Y$.

EXERCISES

Naming of exercises. We will always refer to exercises in Lang by chapter, and then the exercise-number they have in that chapter, preceded by ''$:\mathrm{L}$''. E.g., exercise 3 at the end of Chapter II of Lang is here called ''Exercise II:L3''.

On the other hand, exercises given below are classified by the chapter and section to which they are related. For instance, exercises related to Lang's section VI.2 are numbered VI.2:1, VI.2:2, etc.. (Exercises related to Appendix 2, where Lang gives background on set theory, and to my handout on ''The Axiom of Choice, Zorn's Lemma, and all that'', to be read with that Appendix, are numbered using A2 in place of the chapter number, and are given at the end. The one exercise based on Lang's pp.ix-x, ''Logical Prerequisites'', is numbered 0:1.)

If an exercise is related to more than one section, I will list it under the last section with which it assumes familiarity (using the order of sections noted in the Introduction to this Companion when this differs from Lang's). If one exercise assumes the results of another, I indicate this with the symbol [> ...], meaning that if assigned, it should be given after the other one. In cases where one exercise refers to an idea in another, or perhaps is motivated by a result in another, but doesn't require one to assume that result, I use [cf. ...] rather than [> ...].

Exercises marked * or ¤. When I teach the course, I give homework problems, which all students are expected to do, and a few extra-credit problems, which require more original thinking.

A problem requiring ingenuity can generally be transformed into a straightforward task by adding steps to lead the student through the solution, so the same problem may be given, with modification, as an extra-credit problem one year and a regular homework question the next. Below, I have marked with an asterisk * those exercises that in *the form given here* require considerable original thinking, and could be assigned as extra-credit problems.

It sometimes also happens that an exercise seems difficult until one sees some key trick, which makes it easy. Starting with the 2002 version of these notes, I have begun using another symbol, ''¤'', to mark exercises of this sort. In general, such exercises are not good to assign as homework, but *are* good for students to think about. Until now, I have avoided including such exercises in these notes, so there will be very few until I accumulate more.

(Of course, the starred exercises also generally require some insight or key idea; but in such cases the insight is generally not as straightforward, and still takes some work to apply.)

Notes below on exercises in Lang. Some of the exercises below generalize, clarify, or correct exercises in Lang; notes on some others of Lang's exercises are given at the end of my exercises for each section. For convenience, this collection of exercises ends with an ''index'' to my commentary on Lang's exercises, both in these pages and in the body of this Companion. (But corrections that apply only to earlier printings of Lang are only noted in the Errata, as usual.)

~ ~ ~

0:1. Consider a diagram of sets and set-maps formed by taking the square diagram at the bottom of p.ix, putting a fifth set E in the center, and drawing an arrow connecting E to each of the four objects shown there. (The directions of the arrows will be specified below.)

(a) Show that if the arrow between A and E goes inward (to E) and the arrow between E and D goes outward (to D), then regardless of the direction(s) in which the arrows connecting B and E to E are drawn, commutativity of the four triangles into which those arrows divide the given square implies commutativity of the square.

(This involves four different cases depending on the directions of those arrows; but with a little thought, one can deal with these cases simultaneously.)

(b) Show that if all four arrows point in toward E, or if all four arrows point out from E, then commutativity of those four triangles does not imply commutativity of the square.

I.1:1. Let X be any set, let G be the monoid of all maps $X \to X$, where the ''multiplication'' is composition of maps, and let $H \subseteq G$ denote $\{g \in G \mid \forall x \in X,\ g^{-1}(\{x\})$ is countable$\}$. Show that H is a submonoid of G.

I.1:2. In this exercise, we shall show that there are enormous numbers of isomorphism classes of monoids of small finite cardinalities. It will be easiest to start by considering isomorphism classes of *semigroups*, that is, sets given with an associative composition, not assumed to have a unit. Then, in part (d), we will ''adjoin units''.

(a) Suppose A and B are disjoint sets, c an element not in $A \cup B$, and $f: A \times A \to B$ any set-map. Show that the set $S = A \cup B \cup \{c\}$ becomes a semigroup S_f if we define xy to be $f(x, y)$ if x, y both belong to A, and to be c otherwise. Assuming that A and B are finite, express the number of *distinct* (but not necessarily nonisomorphic) semigroup structures on the set $S = A \cup B \cup \{c\}$ that may be obtained by this construction using various maps f, as a function of the cardinalities $\mathrm{card}(A)$ and $\mathrm{card}(B)$.

(b) Show that given two maps $f, g: A \times A \to B$, any isomorphism between the semigroups S_f and S_g must carry A into A, B into B, and c to c. Deduce that among the semigroup structures described in (a), no more than $\mathrm{card}(A)!\,\mathrm{card}(B)!$ can belong to any one isomorphism class.

(c) Taking $\mathrm{card}(A) = \mathrm{card}(B) = n$, and using the estimate $n! \leq n^n$, obtain a lower bound on the number of isomorphism classes of semigroups of $2n+1$ elements.

(d) Show that if G is any semigroup, and e an element not in G, then $G \cup \{e\}$ can, in a unique manner, be made into a monoid having G as a subsemigroup and e as unit. (Note: do not confuse the uniqueness asked for here with the uniqueness of the identity element of a monoid, proved by Lang on p.3.) Show that if G and H are nonisomorphic semigroups, then $G \cup \{e\}$ and $H \cup \{e\}$ become nonisomorphic monoids.

(e) Deduce a lower bound on the number of isomorphism classes of monoids of $2n+2$ elements.

(f) Find reasonably small integers M and N such that there are at least 100 nonisomorphic monoids of M elements, and at least 10^{100} nonisomorphic monoids of N elements. (You may obtain these bounds using the estimate of point (e), or you may see whether you can get still smaller M and N by ad hoc choices of unequal values for $|A|$ and $|B|$.)

***I.2:1.** Let G be a group and S a subset. Let D be the set of all elements $x \in G$ which have the property that for every group H and every two homomorphisms $p, q: G \to H$ agreeing on S, we also have $p(x) = q(x)$. By the same sort of argument as we gave in the main part of these notes concerning the statement at the bottom of p.10 of Lang, we can see that D is a subgroup of G containing the subgroup generated by S. Must it equal the subgroup generated by S?

I.2:2. (a) Suppose G is a group and H, K are subgroups, and we define $HK = \{hk \mid h \in H, k \in K\}$. Show that HK is a subgroup of G if and only if it is closed under taking inverses.

(b) Show by example that the corresponding statement is false for 3-fold products HJK of subgroups of G.

Note on I:L44 (p.79): The assumption that A and A' be abelian is not needed. This exercise can go with §I.2.

I.2:3. (a) Suppose G and H are groups, and $a, b, c: G \to H$ are *set-maps*, such that for all $x \in G$,

$a(x)\,b(x) = c(x)$. Show that if *any three* of the following conditions hold, so does the fourth:

(i) a is a homomorphism.

(ii) b is a homomorphism.

(iii) c is a homomorphism.

(iv) Every element of $a(G)$ commutes with every element of $b(G)$.

(b) If one takes $H = G$, lets a and b be the identity endomorphism of G, and defines c in the unique way making part (a) applicable, what does part (a) tell us?

(c) If one takes $H = G$, lets a be the identity endomorphism of G, lets c be the trivial endomorphism (taking every element to e), and defines b in the unique way making part (a) applicable, what does part (a) tell us?

I.3:1. = I:L3 (p.75) but with the last sentence made more precise: ''Show that G/G^C is abelian, and that any homomorphism f of G into an abelian group A can be factored $f = hq$, where q is the canonical homomorphism $G \to G/G^C$, and h is a homomorphism of abelian groups, $G/G^C \to A$.'' (Recommendation: Before reading carefully the above revision, think about the version in Lang, and try to guess what it means, since you will frequently encounter such abbreviated statements.)

I.3:2. Do I:L5 (p.75). Show moreover that in this situation, H is the inverse image of the graph in question under the natural map $G \times G' \to G/N \times G'/N'$.

Deduce that if H is *any* subgroup of $G \times G'$, then there exist subgroups $K \subseteq G$, $K' \subseteq G'$ and homomorphisms f, f' from these respective subgroups into a common group L, such that

$$H = \{(k, k') \in K \times K' \mid f(k) = f'(k')\}.$$

Intuitive idea: Regard $H \subseteq G \times G'$ as a relation. If two elements $r_1, r_2 \in G$ are both ''related'' via H to a common element s_1 of G', then since H was assumed a group, $r_1 r_2^{-1}$ is related to e. One finds that the condition of ''relating to a common element'' is an equivalence relation on the group K of elements of G that relate to elements of G'. If one ''glues together'' elements of K via this equivalence relation, and does the analogous gluing job on the group K' of elements of G' that relate to at least one element of G, one gets factor groups K/N and K'/N', and the relation given by H becomes an isomorphism between these, which uniquely determines H.

(Note that for *monoids* in place of groups, the corresponding arguments fail, because of the lack of inverses; and there are submonoids of product monoids which do not arise from isomorphisms of subfactors. An example is the submonoid of $\mathbf{Z} \times \mathbf{Z}$ given by $\{(x, y) \mid x \leq y\}$.)

(Remark on Lang's use of the phrase ''identify ... as'' in I:L5 and elsewhere. This usage is not standard. There is, of course, the standard use of ''to identify X as Y'' meaning to determine that X is Y: ''He identified the specimen as belonging to an extinct species''. And there is the mathematical use of the phrase ''to identify X *with* Y'' meaning to treat X and Y as the same – sometimes by a formal construction such as dividing out by an equivalence relation; other times by ignoring an ''irrelevant'' distinction, for instance, when one identifies S_n with a subgroup of S_{n+1}. It is not clear to me whether Lang is using ''identify as'' in the sense of ''identify with'', or in a blend of the two meanings, e.g., ''we *determine* that N is isomorphic to a normal subgroup of G, and will henceforth *not distinguish* them.'' In any case, I advise you not to adopt this phrase of his.)

I.3:3. (a) Let $A_0 \subseteq A$ and $B_0 \subseteq B$ be subgroups of a group G. Prove the following equalities. (The product-sets named are not assumed to be subgroups.)

$$(A\,B_0) \cap B = (A \cap B)\,B_0, \qquad A \cap (A_0\,B) = A_0\,(A \cap B), \qquad (A\,B_0) \cap (A_0\,B) = A_0\,(A \cap B)\,B_0.$$

(b) Show using the above results (without any further calculations with group elements) that in the diagram for the Butterfly Lemma on p.21 of Lang, the three groups shown immediately above the

unlabeled node have the property that the *intersection* of any two of them is the *product* of the two groups shown immediately below that node. I.e., show that each of the intersections

(1) $(u(U \cap v)) \cap (U \cap V), \qquad (U \cap V) \cap ((u \cap V)\, v), \qquad ((u \cap V)\, v) \cap (u(U \cap v))$

is equal to the product

(2) $(u \cap V)\ (U \cap v).$

Hint: The operation ''$\cap$'' on sets is associative and commutative, so the first expression in (1) is simply the intersection of the sets $u(U \cap v)$, U, and V. An inclusion relation holds between two of these three sets, allowing one of those two to be dropped from the expression. A similar observation applies to the second expression in (1). When these redundant terms have been dropped, each of the expressions in (1) can be transformed using one of the results of part (a).

(c) Verify that the second display on p.21 of Lang is indeed, as he claims, an instance of the third display (= the second boxed equation on p.17), gotten using for ''H'' and ''N'' the groups named on the line after the latter display. Thus, you need to show that H is contained in the normalizer of N, and that the intersection and product sets in the third display equal the appropriate sets in the second display.

I.3:4. Let G be a group and S a subset of G. We claim there is a *least* normal subgroup N of G containing S, i.e., a normal subgroup which contains S, and is contained in every normal subgroup containing S.

There are two natural ways of constructing this subgroup: as an intersection, and as the closure of S under appropriate operations.

(a) Prove the existence of such an N in one of the above ways, and describe precisely the other approach (without going through the proof that it gives the desired group).

(b) Show that if S is closed in G under conjugation (i.e., if for all $s \in S$ and $g \in G$ one has $g s g^{-1} \in S$), then the above subgroup coincides with the subgroup of G generated by S. Deduce from this the normality statement of I:L3.

The group N described above can be called ''the normal subgroup of G generated by S''. (Lang avoids this wording, perhaps feeling that it would suggest that N is generated by S as a group; e.g., in the next-to-last sentence on p.68 he simply calls it the smallest normal subgroup containing S.)

(c) Suppose G is a group, H a subgroup, and S a subset of H. Let N_1 be the normal subgroup of H generated by S, and N_2 the normal subgroup of G generated by S. Prove an inclusion that must hold between N_1 and N_2, but show by example that these subgroups need not be equal, even if H is normal in G.

Note on I:L4 (p.75): This result is proved (without the assumption $K \subset N_H$) as display (c9) of this Companion. (The assumption $K \subset N_H$ has the advantage of guaranteeing that the set HK is a group; but the result is true without it.)

I.4:1. Show that if m and n are positive integers, there is a nontrivial homomorphism from the cyclic group of order m to the cyclic group of order n if and only if m and n are *not* relatively prime. (Hint: To prove concisely that a homomorphism exists from one cyclic group to another which sends the given generator to the desired image, note that Z_m can be described as the factor-group of $\mathbf{Z}$ by the subgroup $m\mathbf{Z}$, and that the *universal property of factor-groups* (triangular diagram on p.14 of Lang) shows just what is required to obtain a homomorphism on a factor-group.)

I.4:2. (a) Show that if a group G is generated by a set S of elements such that every two elements of S commute, then G is abelian.

(b) Deduce that if G is a group with center C, and G/C is cyclic, then G is abelian, i.e., G/C is trivial.

(c) Deduce I:L7 (p.75).

I.4:3. ¤(a) Show that if a group has a largest proper subgroup (i.e., a proper subgroup that contains all other proper subgroups), then it is cyclic of finite order.

(b) For which positive integers n does the cyclic group of order n have a largest proper subgroup?

(c) Show that a group having a smallest nonzero subgroup need not be cyclic. (Hint: Look at the groups described at the bottom of p.9.)

I.4:4. Let's see how far we can generalize the first two parts of the preceding exercise.

¤(a) Show that if a group G has proper subgroups H_1 and H_2, such that every proper subgroup of G is contained either in H_1 or in H_2, then G again must be cyclic of finite order.

(b) For which positive integers n does the cyclic group of order n have the property of part (a)?

Clearly, this above ''pattern'' cannot continue with the ''1'' of the preceding exercise and the ''2'' of part (a) above replaced by arbitrarily large integers, since that would imply that every finite group was cyclic! In fact, things start to change at the next step:

*****(c) Show that if a group G has proper subgroups H_1, H_2 and H_3 such that every proper subgroup of G is contained in one of these, then either G is cyclic of finite order, or G has a homomorphism onto $Z_2 \times Z_2$.

(d) For which positive integers n does the cyclic group of order n have the property assumed in part (c)? Show that the groups having that property also include $Z_2 \times Z_2$ itself, and the two nonabelian groups of order 8 described at the bottom of p.9 of Lang.

(We will continue the investigation of these groups in Exercise I.6:7.)

I.5:1. Let G be a group. A G-set S is called *transitive* if it consists of a single orbit.

(a) Let S and T be transitive G-sets, and s an element of S. Show that S and T are isomorphic as G-sets if and only if there exists $t \in T$ such that $G_s = G_t$.

Note that if H is a subgroup of G, the set of left cosets G/H can be made a G-set under left translation: $gx = \{ga \mid a \in x\}$ $(g \in G,\ x \in G/H)$, and that this G-set will be transitive. Now

(b) Show using (a) that every transitive G-set S is isomorphic to G/H for some subgroup H of G.

(c) For S, T, s as in (a), generalize the result of (a) to a criterion for there to exist a *homomorphism* $S \to T$ as G-sets.

(d) Suppose we are given transitive G-sets S and T, and we know the isotropy subgroups $H = G_s$ and $K = G_t$ of some $s \in S$ and some $t \in T$. Express the criteria of (a) and (c) above in terms of H and K. (Note that we are concerned with criteria for there to be *some* isomorphism or homomorphism between the indicated G-sets; not necessarily one that takes s to t! So, to express the criterion of (a), you need to find a group-theoretic condition on H and K which is equivalent to the statement that there exists a $t' \in T$ with $G_{t'} = G_s$; and you need to adapt similarly the condition you got in (c).)

I.5:2. [> I.5:1] (a) Show that if G is a finitely generated group, then for each nonnegative integer n there are, up to isomorphism, only finitely many n-element G-sets.

(b) Deduce that a finitely generated group has only finitely many subgroups of any finite index n, hence has only countably many subgroups of finite index.

The next two parts show that two weakened versions of the above statement are false.

(c) If X is a set, and G a subgroup of the additive group of all $\mathbf{Z}$-valued functions on X, then for each $x \in X$, let $G(x) = \{ f \in G \mid f(x)$ is even$\}$. Show that each of these groups has finite index in G. Give an example of an *uncountable* set X and a *countable* subgroup G of the above sort, such that all the subgroups $G(x)$ are distinct. (Suggestion: G might be the additive group generated by the characteristic functions of a countable family of subsets of X which separates points, i.e., has the property that for every two distinct points $x, y \in X$ one of these sets contains one of x, y but not the other.) Thus, G is a

countably generated group with uncountably many subgroups of finite index.

(d) Let G be the group of permutations of the integers generated by the translation function $t: n \mapsto n+1$, and the permutation that interchanges 0 and 1 but leaves all other integers fixed. For every set S of *even* integers, let $G(S)$ denote the subgroup of elements of G that fix all members of S. Show that the subgroups $G(S)$ are all distinct. Thus, the finitely generated group G has uncountably many subgroups.

Lest the example of (c) lead us to think that *all* non-finitely-generated groups have uncountably many subgroups, let us note:

(e) Let p be a prime, and let G be the set of rational numbers of the form a/p^n for $a, n \in \mathbf{Z}$. Show that G forms a subgroup of the additive group of $\mathbf{Q}$, that G is not finitely generated, but that G has only countably many subgroups.

I.5:3. (a) Let G be a group, H a subgroup, and S a transitive G-set with an element s such that $G_s = H$. Let $U \subseteq S$ be the fixed-set of H, i.e., the set of $u \in S$ such that $(\forall\, h \in H)\, hu = u$. Show that if H either has finite order or has finite index in G, then $\{g \in G \mid gs \in U\}$ is the normalizer of H in G.

(b) How can the above set of elements be characterized in the general case (when H is not assumed to have finite order or index)?

(c) Show by example that in the general case, the above set need not be the normalizer. (Suggestions: Let G be the permutation group of the integers, and H the subgroup of permutations that fix all negative integers. Or, let G be the group of maps of the real line into itself of the form $x \mapsto ax+b$ for real constants $a \neq 0$ and b, and H the set of such maps such that $a = 1$ and b is an integer.)

I.5:4. (a) = I:L15 (p.76). (Hint: Count the pairs (g, s) such that g fixes s, first as a sum over g, and then as a sum over s.)

(b) Do I:L16 (p.76) without using I:L15. (Hint: G *is* the union of the *left cosets* of H, but these are disjoint. You may ignore the reference to Chapter XIII.)

(c) Show that each of the above results can be deduced easily from the other.

(d) State the lower bounds that your proofs of (a) and (b) give for the number of elements of G having no fixed points, respectively the number of elements not in the union of all conjugates of H.

I.5:5. If G is a group and N a normal subgroup, let $\mathrm{Aut}(G, N)$ denote the group of those $\sigma \in \mathrm{Aut}(G)$ such that $\sigma(N) = N$.

(a) Show that every $\sigma \in \mathrm{Aut}(G, N)$ induces a $\sigma' \in \mathrm{Aut}(G/N)$, and that this construction yields a homomorphism $\mathrm{Aut}(G, N) \to \mathrm{Aut}(G/N)$.

(b) Show by examples that the above homomorphism need not be one-to-one or onto.

(c) We clearly also have a restriction homomorphism $\mathrm{Aut}(G, N) \to \mathrm{Aut}(N)$; combining this with the homomorphism of (a), we get a homomorphism $\mathrm{Aut}(G, N) \to \mathrm{Aut}(N) \times \mathrm{Aut}(G/N)$. Is this always one-to-one? Onto?

I.5:6. Let G be a group of permutations of a set S, and n a positive integer. One says that G is *n-transitive* (or that it acts *n-transitively* on S, or that S is an n-transitive G-set) if S has at least n elements, and if whenever $s_1, \dots, s_n$, $t_1, \dots, t_n$ are elements of S such that $s_1, \dots, s_n$ are distinct and $t_1, \dots, t_n$ are distinct, there exists a $g \in G$ such that $g(s_1) = t_1, \dots, g(s_n) = t_n$.

(a) Show that for all n and m with $m \leq n$, the symmetric group S_n is an m-transitive group of permutations of $\{1, \dots, n\}$.

(b) Given a positive integer n, for what values of m is the alternating group A_n m-transitive as a group of permutations of $\{1, \dots, n\}$?

(c) Show that for p a prime, the group of permutations of the set $\mathbf{Z}/p\mathbf{Z}$ consisting of all maps $x \mapsto$

$ax+b$ ($a,\ b\in \mathbf{Z}/p\mathbf{Z},\ a\neq 0$) is 2-transitive, but not 3-transitive. (This part assumes familiarity with the definition of the ring $\mathbf{Z}/p\mathbf{Z}$ of integers mod p.)

(d), (e), (f) = I:L47 (p.80), parts (a), (b), (c) respectively, excluding the last assertion of (c), which we give separately as:

(g) Show that any 2-transitive group of permutations on a set is *primitive*, in the sense defined in I:L46.

(h) Show by example that the converse to (g) is false.

I.5:7. (a) Show by example that Exercise I:L38(d) (p.79) does not remain true if we delete the condition that n be prime.

(b) Show, however, that again becomes true if we add the condition $s = r+1$.

Of course, (a) and (b) will be subsumed if you do (c), which is not as hard as it might seem.

(c) Determine necessary and sufficient conditions on n, r and s for S_n to be generated by $[123\dots n]$ and $[rs]$.

One may ask, further

(d) Given a transitive subgroup $G\subseteq S_n$, must there be a transposition $[rs]$ which, together with G, generates S_n?

I.5:8. Let G be a group, S a transitive G-set, s an element of S, and $H = G_s$. Let Aut(S) denote the group of *automorphisms* of S (invertible homomorphisms $S\to S$) as a G-set. Prove that Aut(S) $\cong N_H/H$.

Note on I:L57 (p.82, based on §I.5): Both the Examples at the end have the hypothesis $X = GY$ from the preceding paragraph. In Example 2, I think ''discretely'' just means that there is no assumption that the operation is continuous in G. (If so, the word is superfluous, since G was assumed discrete.)

I.6:1. [> I.4:2 and I.5:5(a)]. (a) = I:L24 (p.77). (Suggestions: Get commutativity with the help of Theorem I.6.5 and I.4:2(b) above. If the group has exponent p, show with the help of Proposition I.2.1 that it is isomorphic to $Z_p\times Z_p$, while if it does not have exponent p, show that it is isomorphic to Z_{p^2}.)

(b) Show that if a p-group H acts on a nontrivial p-group G by automorphisms, then its fixed group is nontrivial. (Hint: You hardly need to use the fact that G is a group.)

Now in the situation of (b), if the fixed subgroup is a normal subgroup N, then I.5:5 becomes applicable, and allows you to apply (b) above to G/N as well. This can help you get the remaining two parts:

(c) Suppose G is a group isomorphic to $Z_p\times Z_p$, and θ is an automorphism of G of order p. Show that G is generated by two elements a, b such that $\theta(a) = a$ and $\theta(b) = ab$. Show that this is equivalent to saying that there exists an isomorphism $\alpha\colon G\cong Z_p\times Z_p$ under which θ corresponds to the automorphism φ of $Z_p\times Z_p$ given by $\varphi(i,j) = (i, j+i)$. (Indicate the precise sense in which θ ''corresponds to φ under α''.)

(d) Show that if G is a group isomorphic to Z_{p^2}, the only automorphisms of G of exponent p are the powers of the map $a\mapsto a^{1+p}$.

I.6:2. [> I.6:1]. Let p be a prime. We shall determine here the structures of all noncommutative groups of order p^3.

Suppose G is such a group. By Corollary I.6.6, G has a normal subgroup of order p^2; moreover, by Lemma I.6.7 every subgroup of G of that order is normal. Let us now consider three cases:

(i) G has a normal subgroup N isomorphic to $Z_p\times Z_p$, say generated by elements b and c, and also has an element a of order p outside of N. Deduce that G is a semidirect product of N and the subgroup H generated by a. Show that by choosing generators b and c of N appropriately, one can

get c to commute with a, and to satisfy $a^{-1}ba = bc$. Assuming the generators are so chosen, obtain a precise description of the multiplication of G on elements $a^i b^j c^k$.

Now by I.6:1(a), *either* (i) above holds *or* G has a normal subgroup N isomorphic to $Z_p{}^2$, say with cyclic generator b. (Conceivably, both may be true.) Assuming the latter condition, let a be an element of G not in N. Show that a cannot have order p^3. Thus there are two cases:

(ii) The order of a is p. Deduce that G is again a semidirect product of N and the cyclic group H generated by a, and that this time, by a change of generator of H, i.e., by replacing a if necessary by another generator, one can make $a^{-1}ba = b^{p+1}$. Again, describe the multiplication explicitly.

(iii) The order of a is p^2. Again show that by modifying the choice of a, one can make $a^{-1}ba = b^{p+1}$. Show also that by modifying the choice of b, one can get $b^p = a^p$. Now if we let H denote the subgroup of G generated by a, then the action of H on N by conjugation allows one to construct a semidirect product group of order p^4. Show that G is the quotient of this group by a certain normal subgroup of order p. Verify that every element of G can be written uniquely in the form $a^i b^j$ with $0 \le i < p$, $0 \le j < p^2$, and describe the multiplication of such elements.

Thus, for every p, we have constructed three nonabelian groups of order p^3, such that every group of that order has one of these three forms. Show, now, that if $p = 2$, a certain two of these are isomorphic, if $p \neq 2$, a different two are isomorphic, and that in each case, the remaining one is not isomorphic to the others.

Thus, for each p, there are, up to isomorphism, exactly two nonabelian groups of order p^3.

(From the above facts, it follows that there must be one of constructions (i)-(iii) that we could drop from our list, and then say that for every prime p, every group of order p^3 is isomorphic to exactly one of the two groups on the list. However, for p odd, the construction we would be dropping gives a simpler description of the group in question than the one that is isomorphic to it. So I think this three-constructions-with-isomorphisms description is more enlightening than the two-groups description would be.)

I.6:3. [> 1.5:3(a)]. Except where the contrary is stated (in part (d) below), let G be a p-group.

(a) Suppose H is a proper subgroup of G, and S a transitive G-set with an element s such that $G_s = H$. Show that the set U defined as in I.5:3(a) has cardinality > 1. (Hint: you've had a result about G-sets when G is a p-group.)

(b) Deduce that the normalizer of a proper subgroup H of G is strictly larger than H.

(c) Deduce that any maximal proper subgroup of G is normal.

(d) Show that in *any* group G, a subgroup which is both maximal among all proper subgroups, and normal, must have prime index.

(e) Let G again be a p-group, and let G^p be the subgroup of G generated by all pth powers of elements of G. Show that G^p is normal.

(f) Let G^c be the *commutator subgroup* of G, defined as in Lang, p.20, top paragraph. Show that $G^p G^c$ is the intersection of all maximal proper subgroups of G. (This is called the *Frattini subgroup* of G.)

(g) Show that if S is a subset of G, the following four conditions are equivalent:

(i) S generates G.
(ii) The image of S generates G/G^c.
(iii) The image of S generates G/G^p.
(iv) The image of S generates $G/G^p G^c$.

(Can you generalize the argument by which you got (g) from (f) to a result about groups which are not necessarily p-groups?)

I.6:4. [> I.5:1]. We shall determine all nonabelian groups of order 30.

(a) Prove (by a quick application of a result in the material on ''Structures of groups of order pq'') that every group of order 15 is cyclic (= I:L6).

Now suppose G is a nonabelian group of order 30.

(b) Show by counting elements that G cannot simultaneously have more than one 3-Sylow subgroup and more than one 5-Sylow subgroup.

Thus, G must have either a normal 3-Sylow subgroup or a normal 5-Sylow subgroup.

(c) Suppose G has a normal 3-Sylow (respectively, 5-Sylow) subgroup N. Show that in G/N, a 5-Sylow (respectively, 3-Sylow) subgroup will be normal, and hence that its inverse image in G, which we shall call M, will in turn be normal. Determine the structure of M.

(d) Show that in the above situation, G will be a semidirect product of M with a group of order 2.

(e) Express M as a direct product $M = M_1 \times M_2$, and show that any automorphism of M must carry M_1 into M_1 and M_2 into M_2. Determine all automorphisms of M of exponent 2. Describe all possible structures for G (as explicit semidirect products).

(f) Show that of the groups described above, one and only one is *not* a direct product of two groups of smaller orders. In the remaining cases, describe the direct product decompositions.

I.6:5. (a) If G is a finite group, show that the following conditions are equivalent: (i) G *cannot* be embedded in a direct product of groups of smaller orders. (ii) G has a smallest nontrivial normal subgroup. (Idea: Consider maps $G \to G/N \times G/N'$. Recall that ''smallest'' means ''contained in all others''.)

(b) [> I.6:4] Deduce that the one group G of I.6:4(f) that was not a direct product of groups of smaller orders can be written as a subgroup of a direct product of groups of smaller orders. Describe these groups.

***I.6:6.** Show that for every positive integer k there exists a positive integer N such that every finite group of order $\geq N$ has an abelian subgroup of order $\geq k$. (Suggestion: First do this for p-groups.)

***I.6:7.** (a) Suppose as in I.4:4(c) that G is a group having proper subgroups H_1, H_2 and H_3 such that every proper subgroup of G is contained in one of these. You analyzed in I.4:4(d) the case where G was cyclic. Now show that if G is finite but not cyclic, then it is a 2-group. (Hint: Will a 2-Sylow subgroup of G be proper?)

I do not know whether there are any *infinite* groups satisfying the above assumptions.

***(b)** Show that a finite 2-group has the property of the first sentence of (a) if and only if it can be generated by two, but not by fewer, elements.

(c) Given a group G and a positive integer n, examine the relationships among the three conditions, (i) G has a family of n (not necessarily distinct) proper subgroups, such that every proper subgroup of G is contained in at least one of these. (ii) G has $\leq n$ maximal proper subgroups. (iii) G can be written as a union of $\leq n$ proper subgroups.

In particular, try to determine what implications hold among these three conditions for all n, giving counterexamples to implications that do not always hold. Examine whether any natural restrictions, such as finiteness, finite generation, or non-cyclicity, imply any of the implications that do not hold in general.

Some additional exercises that could be given in connection with §I.6 are indicated in the paragraph preceding Exercise I.9:1.

***I.7:1.** (a) Show that the group $\mathbf{Z}^\omega$ of all sequences $(a_0, a_1, \ldots)$ of integers, under componentwise addition, is *not* free abelian on any generating set.

(b) Can you describe the group $\mathrm{Hom}(\mathbf{Z}^\omega, \mathbf{Z})$?

I.7:2. (a) Show that if M is an abelian group, and we regard it as a commutative monoid and form

$K(M)$, then the canonical map $\gamma\colon M \to K(M)$ is an isomorphism.

(b) On the other hand, show that if M is a commutative monoid containing an element z such that $a + z = z$ for all $a \in M$, then $K(M)$ is the 1-element group. (Such an element z is called a *zero* element of a monoid, because the element 0 has this property in the monoid formed by any ring under multiplication, though the notation there is, of course, $a0 = 0$.)

I.7:3. Let A be a free abelian group, and x an element of A. Show that x is a member of a basis of A if and only if it cannot be written ny for $n > 1$ and $y \in A$.

(Hint for ''if'': Suppose x is written $\Sigma_{i \in I}\, n_i\, e_i$ where $\{e_i \mid i \in I\}$ is a basis of A, and the n_i are integers. Consider the subgroup of $\mathbf{Z}$ generated by $\{n_i \mid i \in I\}$.)

¤**I.7:4.** Find an example of an abelian group A which is the sum, $A = B + C$, of two *free* abelian subgroups, but which is not itself free abelian.

Of course, this means, necessarily, that this sum is not a direct sum. However, can you do this in such a way that B and/or C is a direct summand in A? (The condition that B is a direct summand means that there exists a subgroup B' such that $A = B \oplus B'$; and similarly for C.) In such a way that A is torsion-free?

Note: Exercises III:L5-III:L8 can go with §I.7, assuming a little background knowledge: For III:L5 one needs the concept of a basis of n-dimensional real Euclidean space $\mathbf{R}^n$, and for III:L8, the arithmetic of the rational numbers, and the notation $\mathbf{Q}^*$ for the multiplicative group of nonzero rationals. Some corrections to III:L6, are noted after the exercises going with §III.5 below.

I.8:1. Let G be the group of all sequences $(a_0, a_1, a_2, \ldots)$ with $a_i \in Z_4$, under componentwise addition, and let H be the subgroup of G consisting of those elements such that $2a_0 = 0$. Show that each of G and H can be embedded in the other, but that G and H are not isomorphic.

I.8:2. Suppose G is a group such that every finitely generated subgroup of G is cyclic.

(a) Show that if G is torsion-free, then G can be embedded in the additive group $\mathbf{Q}$ of rational numbers.

***(b)** Show that if G has torsion, then G can be embedded in the factor-group $\mathbf{Q}/\mathbf{Z}$.

***I.8:3.** Suppose p is a prime, and define $Z_{p^\infty} = (\mathbf{Q}/\mathbf{Z})(p)$, with the ''$(p)$'' notation as in Lang (p.42, paragraph before Theorem I.8.1). Show that $\mathrm{Aut}(Z_{p^\infty})$ is uncountable.

I.8:4. [> A2.2:4(a), at end of these exercises].

(a) Show that the subdirectly irreducible finitely generated abelian groups are precisely the cyclic groups of prime-power order.

(b) Show that the subdirectly irreducible finite p-groups are the p-groups whose centers are cyclic.

(c) [> I.8:2] Assuming the results of both parts of I.8:2, show that the subdirectly irreducible abelian groups that are not finite are the groups Z_{p^∞} defined in I.8:3 above.

Remark on the next five exercises. We know from results in Lang that all finitely generated torsion abelian groups are direct sums of finite cyclic groups Z_{p^n}, and we may ask whether something like this is true without the finite generation hypothesis. The next four exercises address this question. Since the groups Z_{p^∞} defined in Exercise I.8:3 above are not direct sums of proper subgroups, any positive result along these lines will either have to allow these groups as summands along with the cyclic groups, or have some hypothesis which excludes them.

Exercise I.8:5 will show that even allowing summands Z_{p^∞}, we do not get such a direct sum decomposition for arbitrary torsion abelian groups. On the other hand, in Theorem I.8.c2 in this Companion we got a generalization of the abovementioned result from Lang to the non-finitely-generated

case by strengthening the hypothesis of being torsion to a condition that in the finitely generated case is equivalent to being torsion. Two other proofs of that same result are indicated in Exercises I.8:6 and I.8:7 below. Exercise I.8:8, on the other hand, will indicate a result for arbitrary torsion abelian groups, obtained by weakening the conclusion of being a direct sum of cyclic subgroups to one that is equivalent, in the finitely generated case, to that conclusion.

Exercise I.8:9 shows that yet another result true in the finitely generated case fails for non-finitely-generated groups.

***I.8:5.** Let p be a prime, and A the torsion subgroup of $\prod_{1 \leq i < \infty} Z_{p^i}$. Show that A contains no subgroup isomorphic to Z_{p^∞}, and is not a direct sum of cyclic subgroups.

I.8:6. (Alternative proof of Theorem I.8.c2, this Companion.)

In (a)-(c), let A be an abelian group of prime-power exponent p^n. We shall write A_p for $\{a \in A \mid pa = 0\}$, and for each $i \geq 0$ we shall write $p^i A$ for the subgroup $\{p^i a \mid a \in A\} \subseteq A$.

(a) Show that one can construct, successively, subsets S_n, S_{n-1}, ..., S_1 of A_p, such that for each i, S_i is a subset of $A_p \cap p^{i-1}A$ maximal for the condition that no $s \in S_i$ lies in the subgroup of A generated by $S_n \cup S_{n-1} \cup \ldots \cup S_i - \{s\}$.

In (b) and (c) below, $S_1, \ldots, S_n$ denote sets with this property.

(b) Show that A_p is the direct sum of its cyclic subgroups $\langle s \rangle$ for $s \in \bigcup S_i$.

(c) Suppose we choose, for each element s of each S_i, an element $t \in A$ such that $p^{i-1}t = s$. (Why is this possible?) Letting T_1, ..., T_n be the sets of elements so obtained, and $T = \bigcup T_i$, show that A is the direct sum of the cyclic subgroups $\langle t \rangle$ $(t \in T)$.

(d) Deduce that every abelian group of finite exponent n is a direct sum of cyclic subgroups.

***I.8:7.** (Yet another proof of Theorem I.8.c2, with most of the work left to you this time.)

Suppose p is a prime, and A an abelian group all of whose elements have exponents a power of p. Let us call a subgroup $B \subseteq A$ *pure* if $B \cap pA = pB$.

(a) Show that the least integer i such that A has a nonzero pure subgroup of exponent p^i is the same as the least integer i such that A has a maximal cyclic subgroup of exponent p^i.

(b) Show that for i as in part (a), any pure subgroup $B \subseteq A$ of exponent p^i is a direct summand. (Hint: Show that $p^i A \cap B = \{0\}$, and that $p^i A$ can be enlarged to the desired complementary summand.)

(c) Deduce that every abelian group of finite exponent r is a direct sum of cyclic subgroups.

***I.8:8.** [> I.8:7]. Let A be an abelian p-group. Show that, up to isomorphism, one can write

$$\bigoplus_{1 \leq i \leq \infty} A_i \;\subseteq\; A \;\subseteq\; \prod_{1 \leq i \leq \infty} A_i,$$

where for each i, A_i is a direct sum of copies of Z_{p^i}. In the case $i = \infty$, Z_{p^∞} is defined as in I.8:3.

(Idea: Get decompositions $A = A_1 \oplus B_1$, $B_1 = A_2 \oplus B_2$, etc., and show that such a chain of decompositions leads to the situation of the above display, where A_∞ is the intersection of the B_i.)

I.8:9. Let A be an abelian group.

We have seen that if A is finitely generated, then it is the direct sum of its torsion subgroup A_{tor}, and a free abelian subgroup B. In the non-finitely-generated case, a torsion-free abelian group need not be *free* abelian, but we can still ask whether A will be the direct sum of A_{tor} and a torsion-free subgroup B. In (b) and (c) below we shall see two examples showing that this is not so in general (with a modified argument for case (c) given in (d)). The criterion we shall use in both (b) and (c) is proved in (a). In (e)-(f) we shall examine further properties of these examples.

(a) Show that if C is any subgroup of A, then it is a direct summand in A (i.e., there is a subgroup B of A such that $A = B \oplus C$) if and only if the canonical surjection $A \to A/C$ has a right inverse which

is a group homomorphism, and that in this case, we will have $B \cong A/C$.

In the two examples below, you will be able to prove that A_{tor} is not a direct summand by showing that A/A_{tor} has properties that no subgroup of A has. To state these properties, let us say that an element a of an abelian group (written additively) is ''divisible by'' an integer n if $a = nx$ for some group element x. Of the two examples, the first will be easier to state, but the second perhaps gives better insight into why a direct sum decomposition is impossible.

(b) Let A be the direct product, over all primes p, of the cyclic groups Z_p. Show that every element of A/A_{tor} is divisible by *all* positive integers n, but that no nonzero member of A has this property. Deduce that A_{tor} is not a direct summand in A.

(c) Let p be any prime, and let $\mathbf{Z}[p^{-1}]$ denote the abelian group of all rational numbers that can be written with denominator a power of p. (This is ring-theoretic notation for ''the ring generated by adjoining p^{-1} to the integers''; but here we are using it as shorthand for ''the underlying additive abelian group of that ring''.) Let us form the factor-group $\mathbf{Z}[p^{-1}]/\mathbf{Z}$, and for each positive integer n, let $(p^{-n}\mathbf{Z})/\mathbf{Z}$ denote the cyclic subgroup thereof generated by $[p^{-n}]$ (the image in this factor-group of $p^{-n} \in \mathbf{Z}[p^{-1}]$).

The group $\mathbf{Z}[p^{-1}]/\mathbf{Z}$ is what we have previously called Z_{p^∞}; we are describing it this way because we will want to use the canonical map $\mathbf{Z}[p^{-1}] \to \mathbf{Z}[p^{-1}]/\mathbf{Z}$, which we will write $a \mapsto [a]$. Indeed, let us define the subgroup

$$A \subseteq \mathbf{Z}[p^{-1}] \oplus (p^{-1}\mathbf{Z})/\mathbf{Z} \oplus (p^{-2}\mathbf{Z})/\mathbf{Z} \oplus (p^{-3}\mathbf{Z})/\mathbf{Z} \oplus \ldots .$$

to consist of all elements $(a, b_1, b_2, b_3, \ldots)$ such that $[a] = b_1 + b_2 + b_3 + \ldots$. Show that $A/A_{\mathrm{tor}} \cong \mathbf{Z}[p^{-1}]$, so that, in particular, every element of this factor-group is divisible by all powers of p, but that no element of A which maps to a noninteger element of $\mathbf{Z}[p^{-1}]$ is divisible by all powers of p. Deduce that A_{tor} is not a direct summand in A.

(d) In example (c), determine all subgroups $B \subseteq A$ maximal for the property of being disjoint from A_{tor}, and show directly that none of these satisfies $A = B \oplus A_{\mathrm{tor}}$.

We now look at a few related properties of these examples.

(e) In example (b), show that for every prime p, the p-torsion subgroup $A(p) \subseteq A$ *is* a direct summand in A; and in fact that there is a *unique* subgroup B_p such that $A = A(p) \oplus B_p$. What is the intersection of all the subgroups B_p?

(f) In example (c), let $x_0 \in A$ be the element $([1], 0, 0, 0, \ldots)$. Show that x_0 is divisible by p^n for all n, but that there do not exist elements $x_1, x_2, \ldots$ such that $x_0 = px_1$, $x_1 = px_2$, $\ldots$, $x_i = px_{i+1}$, $\ldots$.

Remark: The divisibility arguments of (b) and (c) above gave relatively simple ways of showing that A/A_{tor} was not embeddable in A. But one can also get examples in which the one-variable equations $a = nx$ defining divisibility are replaced by systems of several linear equations in several variables. In this way one can get examples where A/A_{tor} has no nonzero elements divisible by infinitely many integers, but A_{tor} is still not embeddable in A.

The following three exercises, though I am putting them under §I.9, could be given any time after §I.2, referring to §I.9 only for the definition of *bilinear map* in the middle of p.48. The first of these exercises goes naturally with §I.9, but the next two are appropriate any time after §I.6 has been covered.

I.9:1. If A and B are abelian groups, we note that the set of group homomorphisms $A \to B$ becomes an abelian group under the operation defined by $(f+g)(a) = f(a) + g(a)$. Let us denote this group $\mathrm{Hom}(A, B)$. For three abelian groups A, B and C, the set of *bilinear maps* $A \times B \to C$ may, in the same way, be made a group $\mathrm{Bil}(A, B, C)$. (Take these two facts for granted in your write-up.)

Show that given abelian groups A, B, C there exists an isomorphism θ: $\mathrm{Bil}(A, B, C) \to \mathrm{Hom}(A, \mathrm{Hom}(B, C))$ characterized by the formula $\theta(\beta)(a)(b) = \beta(a, b)$.

I.9:2. (a) Suppose A and B are abelian groups (written additively), and $\beta\colon A\times A \to B$ any bilinear map. Show that the set $A\times B$ becomes a group G_β, in general nonabelian, under the operation $(x,y)(x',y') = (x+x',\ y+y'+\beta(x,x'))$.

(b) For any two elements x, y of a group G, let us write $[x,y]$ for the commutator $xyx^{-1}y^{-1}$. (The commonest symbol for $xyx^{-1}y^{-1}$ is (x,y), but since we are working with ordered pairs $(x,y)\in A\times B$, I felt that the use of that notation for commutators could lead to confusion.) Show that the groups G_β constructed in (a) all satisfy the identity $[[x,y],z] = e$.

(c) Suppose the abelian groups A and B of part (a) both have prime exponent p. Show that if p is odd, then G_β will again have exponent p, while if $p = 2$, G_β may not have exponent 2, but will have exponent 4.

(d) Deduce from (a) that if A, B, C are three abelian groups, and $\beta\colon A\times B \to C$ a bilinear map, then the set $A\times B\times C$ becomes a group H_β under the operation

$$(x,y,z)\ (x',y',z') \ = \ (x+x',\ y+y',\ z+z'+\beta(x',y)) \qquad (x,\ x'\in A,\ \ y,\ y'\in B,\ \ z,\ z'\in C).$$

Deduce that these groups also have the properties stated in (b) and (c).

(e) Show that the group H_β described in (d) above can be written as a semidirect product $A\leftthreetimes(B\times C)$ and as a semidirect product $B\leftthreetimes(A\times C)$. (If your proof is not based on an explicit description of the maps $A \to \mathrm{Aut}(B\times C)$ and $B \to \mathrm{Aut}(A\times C)$ used in these semidirect product constructions, and the bijections between the set of 3-tuples which form H_β and the sets of pairs which form these semidirect product groups, you should determine retrospectively what these maps and bijections are. Since expressing H_β as a semidirect product $B\leftthreetimes(A\times C)$ involves changing the order of the factors A and B, and elements of those factors do not commute with each other, the above bijection can be expected to be nontrivial in this case.)

(f) On the other hand, give an example of a group of the form G_β as in (a), with A and B both nontrivial, which is not a semidirect product of proper subgroups.

I.9:3. [> I.9:2(a), and (d) excluding last sentence]. In this exercise we will prove that for large n, there are ''many'' groups of order p^n, by a method similar to that used for monoids in Exercise I.1:2, but with bilinear maps of finite products of Z_p in place of set-maps of finite sets. Hence let us begin by counting bilinear maps.

(The results you are asked to prove below are actually true without the assumption that p is prime; I have included that assumption because p-groups form a natural interesting class of groups.)

(a) Suppose a, b and c are positive integers, and p a prime. In the abelian group $(Z_p)^a$, let $x_1, \dots, x_a$ denote the elements $(1,0,\dots,0)$, $\dots$, $(0,0,\dots,1)$; the analogous families of elements of $(Z_p)^b$ and $(Z_p)^c$ will be denoted $y_1,\dots,y_b$ and $z_1,\dots,z_c$. Show that every bilinear map $\beta\colon (Z_p)^a\times(Z_p)^b \to (Z_p)^c$ is determined by the ab elements $\beta(x_i,\ y_j)\in(Z_p)^c$, and conversely, that each ab-tuple of values in $(Z_p)^c$ can be realized as the images of this family of elements under a bilinear map. Thus, determine the precise number of bilinear maps $(Z_p)^a\times(Z_p)^b \to (Z_p)^c$. (We don't really need p to be prime here; I am just assuming this because we will want it later.)

Now by I.9:2(d), each such bilinear map β determines a group H_β with underlying set $(Z_p)^a\times(Z_p)^b\times(Z_p)^c$ (which we may think of as $(Z_p)^{a+b+c}$). To complete our argument, we need to show that not too many of these are isomorphic. It would be nice if we could say that any isomorphism between two such groups was induced by automorphisms of $(Z_p)^a$, $(Z_p)^b$ and $(Z_p)^c$, but this is not so. Rather, our trick will be to introduce some additional structure, which allows us to distinguish among the groups determined by any two bilinear maps, then show that taking away this structure does not cause too many groups to fall together.

Hence, if n is a positive integer, let an *n-pointed group* mean a system $(G,\ u_1,\dots,u_n)$, where G is a group, and $u_1,\dots,u_n$ are elements of G. A homomorphism of n-pointed groups $(G,\ u_1,\dots,u_n) \to$

$(H, v_1, \dots, v_n)$ will mean a homomorphism of underlying groups that takes u_i to v_i for $i = 1, \dots, n$. Now if $\beta\colon (Z_p)^a \times (Z_p)^b \to (Z_p)^c$ is a bilinear map, let H'_β denote the $a+b+c$-pointed group whose underlying group is H_β, defined as in Exercise I.9:2(d), and whose $a+b+c$ distinguished elements are $(x_i, 0, 0)$ $(1 \le i \le a)$, $(0, y_j, 0)$ $(1 \le j \le b)$, and $(0, 0, z_k)$ $(1 \le k \le c)$ (or, regarding the underlying set of H'_β as $(Z_p)^{a+b+c}$, the $a+b+c$ elements $(1, 0, \dots, 0)$, ... , $(0, 0, \dots, 1)$).

(b) Show that if a, b, c are positive integers, and β, γ are bilinear maps $(Z_p)^a \times (Z_p)^b \to (Z_p)^c$, then there exists a homomorphism $H'_\beta \to H'_\gamma$ as $a+b+c$-pointed groups if and only if $\beta = \gamma$, and that in this case, the only such homomorphism is the identity. (Warning: One may be tempted to call on the result at the bottom of p.10 of Lang, and argue that a homomorphism $H'_\beta \to H'_\gamma$ which agrees with the identity map of underlying sets on a generating set for H'_β must agree with the identity map on all elements. This is not valid, because that statement refers to uniqueness of *homomorphisms*, and till we prove what we are trying to prove, we cannot say that the identity map of underlying sets is a homomorphism $H'_\beta \to H'_\gamma$. So you need to do some calculating.) Deduce a lower bound on the number of nonisomorphic $a+b+c$-pointed groups of order p^{a+b+c}, as a function of p, a, b and c.

(c) Determine the number of $a+b+c$-pointed group structures on a group of order p^{a+b+c}. Comparing this result with those of (a) and (b), get a lower bound on the number of nonisomorphic groups of order p^{a+b+c}.

(d) Find reasonably small integers M and N such that the above result implies that the number of isomorphism classes of group of order p^M approaches infinity as $p \to \infty$, and that the number of isomorphism classes of group of order p^N is at least p^{1000}.

I.9:4. (a) Suppose A and B are abelian groups, and A_0, A_1 are subgroups of A and $f_0\colon A_0 \to B$, $f_1\colon A_1 \to B$ homomorphisms, such that for all $x \in A_0 \cap A_1$ one has $f_0(x) = f_1(x)$. Show that a homomorphism $f\colon A_0 + A_1 \to B$ may be defined by setting $f(x_0 + x_1) = f_0(x_0) + f_1(x_1)$.

In parts (b) through (e) below, m will be a fixed positive integer, dual groups will be defined with respect to a fixed cyclic group Z_m, and A and B, where they occur, will denote abelian groups of exponent m.

(b) Let A' be a subgroup of A, and let $\varphi \in A'^\wedge$. Show that there exists an element $\psi \in A^\wedge$ whose restriction to A' is φ. (Hint: If you have extended φ to a subgroup $A_0 \supseteq A$, and if a is an element not in A_0, let $A_1 = \langle a \rangle$, and set things up so that part (a) can be applied.)

(c) Show that $A \neq 0 \Rightarrow A^\wedge \neq 0$.

(d) If $f\colon A \to B$ is a homomorphism, show that f is surjective if and only if $f^\wedge$ is injective.

(e) If $f\colon A \to B$ is a homomorphism, show that f is injective if and only if $f^\wedge$ is surjective.

(Remark: If A and B are finite, then in view of Proposition I.9.c1, part (c) above is trivial, either of (d), (e) is immediate from the other, and even (a) is fairly easy. But for general A and B, these shortcuts are not available.)

(f) Show by example that the result of (a) fails if A and B are not assumed abelian. (Does it become true if one or the other of those groups is assumed abelian?)

I.9:5. Show by example that if, in Theorem I.9.4, we delete the assumption that A'/B' is finite, then the conclusion that $(A'/B') \cong (A/B)^\wedge$ can be false.

I.9:6. Let $(A_i)_{i \in I}$ be a family (in general infinite) of groups of exponent m.

(a) Show that $(\bigoplus A_i)^\wedge \cong \prod (A_i^\wedge)$.

(b) Show by example that the relation $(\prod A_i)^\wedge \cong \bigoplus (A_i^\wedge)$ does not in general hold.

(c) Show that of the two groups named in part (b), one can always be embedded in the other. (In fact, you should be able to describe a map that always gives such an embedding.)

I.11:1. Let $f\colon A \to B$ be a morphism in a category $\mathbf{C}$. Then f is called a *monomorphism* if for every object X of $\mathbf{C}$ and every pair of morphisms $i, j\colon X \to A$, we have $i \neq j \Rightarrow fi \neq fj$. Dually, f is called an *epimorphism* if for every object Y and pair of morphisms $i, j\colon B \to Y$, we have $i \neq j \Rightarrow if \neq jf$.

(a) Show that in **Set**, the monomorphisms are the one-to-one set maps and the epimorphisms are the surjective set-maps.

(b) Show that in **Monoid**, the monomorphisms are the one-to-one homomorphisms, and the class of epimorphisms includes the surjective homomorphisms.

(c) Show that the inclusion-map of the monoid of nonnegative integers under addition into the monoid of all integers under addition is an epimorphism in **Monoid**, though it is not surjective.

(d) Show that in any category $\mathbf{C}$, if $p\colon X \to Y$ and $q\colon Y \to X$ are morphisms such that $pq = \mathrm{id}_Y$, then p is an epimorphism and q a monomorphism.

I.11:2. Let $\mathbf{A}$ and $\mathbf{B}$ be categories. Show that we can define a category $\mathbf{B}^{\mathbf{A}}$ whose objects are the functors $\mathbf{A} \to \mathbf{B}$, and whose morphisms are the morphisms of functors (as defined on p.65). More precisely, show that when we define composition of morphisms and identity morphisms of functors in the obvious ways, the definition of a category is satisfied except for the axiom CAT1, and indicate briefly how to ''fix'' this deficiency.

I.11:3. [> I.11:2]. Let $\cdot\!\to\!\cdot$ denote the category having two objects, named 0 and 1, and three morphisms: id_0, id_1 and a morphism $\varphi\colon 0 \to 1$. (You should verify for yourself that these conditions determine an essentially unique category, but should not hand in the verification.) Thus, following the notation of the preceding exercise, $\mathbf{Group}^{\cdot\to\cdot}$ denotes the category whose objects are the functors from $\cdot\!\to\!\cdot$ to **Group**, and whose morphisms are the morphisms of functors.

(a) Describe in group-theoretic terms the data that determine an object of $\mathbf{Group}^{\cdot\to\cdot}$ (without referring to $\cdot\!\to\!\cdot$, or to the concept of functor), and what constitutes a morphism between two such objects.

(b) Show that the constructions associating to every group homomorphism $f\colon G \to H$ the group $\mathrm{Ker}(f) \subseteq G$ and the group $\mathrm{Im}(f) \subseteq H$ can be regarded as functors $\mathbf{Group}^{\cdot\to\cdot} \to \mathbf{Group}$.

I.11:4. Suppose $\mathbf{C}$ is a category such that any two objects X and Y of $\mathbf{C}$ have a product $X \times Y$ in $\mathbf{C}$.

(a) Show that any three objects X, Y, Z of $\mathbf{C}$ must then have a product, given by $(X \times Y) \times Z$.

(b) In this situation, show that $(X \times Y) \times Z \cong X \times (Y \times Z)$.

I.11:5. In a category $\mathbf{C}$, let us define a *retract* of an object X to mean an object Y given with morphisms $f\colon X \to Y$, $g\colon Y \to X$ such that $fg = \mathrm{id}_Y$; or, more precisely, a 3-tuple (Y, f, g) such that Y, f and g satisfy these conditions.

(a) Show that if (Y_1, f_1, g_1) and (Y_2, f_2, g_2) are retracts of the same object X, and if $g_1 f_1 = g_2 f_2$, then $Y_1 \cong Y_2$ in $\mathbf{C}$.

(b) Show, in fact, that the isomorphism constructed in part (a) will make commuting triangles with f_1 and f_2, and with g_1 and g_2.

(c) Show, conversely, that if two retracts (Y_1, f_1, g_1) and (Y_2, f_2, g_2) of an object X admit isomorphisms making such commuting triangles, then $g_1 f_1 = g_2 f_2$.

(d) Show that there exists a category $\mathbf{R}$ such that for every category $\mathbf{C}$, the set of functors $\mathbf{R} \to \mathbf{C}$ is in natural bijective correspondence with the set of families X, Y, f, g such that (Y, f, g) is a retract of X.

I.11:6. (a) Show that if G is any monoid, then one can construct a category $\mathbf{C}_G$ by taking for object-set a singleton $\{X\}$, and defining $\mathrm{Mor}(X, X) = G$, with the same composition as in the monoid G.

For the next two parts, define an ''action'' of a monoid G on an object Y of any category $\mathbf{C}$ to mean a monoid homomorphism $G \to \mathrm{Mor}(Y, Y)$.

(b) Verify that if G is a *group* and $\mathbf{C} = \mathbf{Set}$, this is equivalent to Lang's definition of an action of G on Y (§I.5).

(c) Show that if G is a monoid and $\mathbf{C}$ a category, then an object of $\mathbf{C}$ with an action of G on it is equivalent to a (covariant) functor from $\mathbf{C}_G$ (defined in (a)) to $\mathbf{C}$.

I.11:7. Let m be a positive integer, and let us write $\mathbf{FinAb}(m)$ for the category of finite abelian groups of exponent m. Show that the duality $h^{Z_m}: A \mapsto A^\wedge$ of §I.9, regarded as a contravariant functor $\mathbf{FinAb}(m) \to \mathbf{FinAb}(m)$, takes short exact sequences to short exact sequences. (Cf. Corollary I.9.2.) (Note our slight abuse of notation: strictly speaking, h^{Z_m} should denote a functor to **Set**; but we are extending this symbol to mean the functor to $\mathbf{FinAb}(m)$ gotten by putting the natural abelian group structures on these hom-sets.)

I.12:1. Let $F = \langle \ldots, y_{-1}, y_0, y_1, \ldots \rangle$ be the free group on a denumerable set of generators $\{y_i \mid i \in \mathbf{Z}\}$, and let $\langle x \rangle$ be an infinite cyclic group (the free group on one generator).

(a) Show that there exists a unique homomorphism $\psi: \langle x \rangle \to \mathrm{Aut}(F)$ such that $\psi(x)(y_i) = y_{i+1}$ for each $i \in \mathbf{Z}$.

(b) Show that the semidirect product $\langle x \rangle \ltimes_\psi F$ is *free* on the two elements x, y_0.

Remark: both parts, (a) and (b) above, can be done via messy calculations, or elegantly, with the help of the universal property of free groups!

I.12:2. (a) Show that the following four groups are isomorphic. (It is up to you to discover the isomorphisms. I said in the discussion of semidirect products that the group described in (iii) is called the ''infinite dihedral group''. Actually, since the groups described below are isomorphic, that term may be defined in any of these four ways.)

(i) The group of maps $T_{a,b}: \mathbf{Z} \to \mathbf{Z}$, where $T_{a,b}(n) = an + b$, for $a \in \{\pm 1\}$, $b \in \mathbf{Z}$. (This is the set of all *distance-preserving* set-maps $\mathbf{Z} \to \mathbf{Z}$, but I don't ask you to prove that.)

(ii) The group with presentation $\langle x, y \mid x^2 = e,\ yx = xy^{-1} \rangle$.

(iii) The group $Z_2 \ltimes_\psi \mathbf{Z}$, where $\psi: Z_2 \to \mathrm{Aut}(\mathbf{Z})$ sends the nonidentity element of Z_2 to the automorphism $n \mapsto -n$ of $\mathbf{Z}$.

(iv) The coproduct of two copies of the group Z_2. (Write these two copies $\{e, u\}$ and $\{e, v\}$, for uniformity of notation.)

(b) Show that in the group (iv) above, every cyclic subgroup generated by an element of the form $(uv)^c$ is normal, and that with three exceptions, every normal subgroup of this group has this form.

I.12:3. Let F be the free group on a denumerable set $\{x_1, \ldots, x_n, \ldots\}$. Given an element $w \in F$, a group G is said to *satisfy the identity* $w = e$ if every homomorphism $F \to G$ carries w to e. A class $\mathbf{V}$ of groups is called a *variety of groups* if there exists some $S \subseteq F$ such that $\mathbf{V}$ is the class of groups satisfying the identities $w = e$ for all $w \in S$.

(a) Verify that the following classes of groups are varieties: (i) the class of all groups, (ii) the class of all abelian groups, (iii) the class of all groups of exponent 50, (iv) the class of all abelian groups of exponent 50, (v) the class of all trivial groups (i.e., up to isomorphism, ''the'' trivial group).

In *one* of cases (ii)-(iv) above, show explicitly how the definition given above of what it means for a group to satisfy an identity (via homomorphisms $F \to G$), when applied to the identities you give, is equivalent to the ordinary definition of the class of groups named. In the remaining cases, you may simply display an S which ''obviously'' works in the same way.

(b) Show that for each positive integer n, the class $\mathbf{Solv}_n$ of all groups G admitting a normal tower of height n with abelian factors:

$$G = G_0 \rhd \ldots \rhd G_n = \{e\} \qquad \text{with } G_i/G_{i+1} \text{ abelian,}$$

is a variety.

(c) Suppose X is *any* set, $F(X)$ the free group on X, and $w \in F(X)$. Let us define the condition that a group G ''satisfies the identity $w = e$'' exactly as in the case $X = \{x_1, \dots, x_n, \dots\}$ above. Show that for every $w \in F(X)$ there exists some $w' \in F(\{x_1, \dots, x_n, \dots\})$ such that a group G satisfies the identity $w = e$ if and only if it satisfies the identity $w' = e$.

(d) Let $w \in F$ (where F is again the free group on $\{x_1, \dots, x_i, \dots\}$). Show that the following conditions are equivalent: (i) The variety determined by the identity $w = e$ consists (up to isomorphism) of the trivial group only. (ii) There exists a group homomorphism $F \to \mathbf{Z}$ taking w to 1. (iii) If we denote by d_i the ''exponent-sum'' of x_i in w, i.e., the sum of all exponents to which x_i occurs, counting signs, in the expression for w, then the greatest common divisor of $\{d_1, \dots, d_i, \dots\}$ is 1.

(This shows that if you write down an identity at random, it is likely not to define an interesting variety.)

I.12:4. [> I.12:3(a)]. (a) Show that for every class $\mathbf{C}$ of groups, there exists a *least* variety $\mathbf{V}$ of groups containing $\mathbf{C}$.

The next four parts will show how to construct the groups in $\mathbf{V}$ from those in $\mathbf{C}$.

If $\mathbf{C}$ is a class of groups, let us define $\mathbf{H}(\mathbf{C})$ to mean the class of *homomorphic images* of groups in $\mathbf{C}$, $\mathbf{S}(\mathbf{C})$ to mean the class of groups embeddable in groups in $\mathbf{C}$ (i.e., groups isomorphic to *subgroups* of members of $\mathbf{C}$) and $\mathbf{P}(\mathbf{C})$ to mean the class of groups isomorphic to *products* $\prod_I G_i$ of arbitrary families $(G_i)_{i \in I}$ of groups $G_i \in \mathbf{C}$ (possibly with repeated factors).

(b) Show that every variety $\mathbf{V}$ of groups is closed under $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{P}$ (i.e., that $\mathbf{V} = \mathbf{H}(\mathbf{V}) = \mathbf{S}(\mathbf{V}) = \mathbf{P}(\mathbf{V})$).

(c) Show that if $\mathbf{C}$ is any class of groups, then the class $\mathbf{HSP}(\mathbf{C})$ is closed under $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{P}$.

(d) If $\mathbf{C}$ is a class of groups, R a member of $\mathbf{C}$, X a set, and $f: X \to R$ a map, let us say that R is ''free on X relative to $\mathbf{C}$'' if the condition defining a free group (Lang, p.66) is satisfied with ''group'' replaced by ''member of $\mathbf{C}$'' throughout. Show that if $\mathbf{C}$ is any class of groups closed under $\mathbf{S}$ and $\mathbf{P}$, then the method of proof of Proposition I.12.1 may be generalized to show that for every set X there exists a group R in $\mathbf{C}$ that is free on X relative to $\mathbf{C}$. Do not repeat that proof – just indicate what observations are used in adapting it.

(As noted in the discussion in this Companion regarding p.58 of Lang, there are various ways of talking about ''free groups'': as pairs (F, f), as universal maps f alone, or as objects F with the map f thought of as ''given''. In parts (d)-(f) of this exercise, I am following the last of these.)

(e) Let X be a set, $\mathbf{V}$ a variety of groups, $F(X)$ the free group on X, and $F_{\mathbf{V}}(X)$ the free group on X relative to $\mathbf{V}$. Show that there is a canonical surjective homomorphism $q: F(X) \to F_{\mathbf{V}}(X)$, and show that for any $w \in F(X)$, the identity $w = e$ holds in $\mathbf{V}$ if and only if $q(w) = e$ in $F_{\mathbf{V}}(X)$.

(f) (Birkhoff's Theorem, for the case of groups) Show with the help of the preceding results that for any class $\mathbf{C}$ of groups, $\mathbf{HSP}(\mathbf{C})$ is the least *variety of groups* containing $\mathbf{C}$. (Hint: We did not use ''$\mathbf{H}$'' in part (d).) Deduce in turn that a class of groups is a variety if and only if it is closed under $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{P}$.

I.12:5. Show that the *free abelian group* on generators $x_1, \dots, x_n$ may be presented *as a group* by the n generators $x_1, \dots, x_n$, and the $n(n-1)/2$ relations $x_i x_j = x_j x_i$ $(1 \le i < j \le n)$.

***I.12:6.** [cf. I.12:5.] Show that no presentation of the free abelian group on n generators *as a group* can involve fewer than $n(n-1)/2$ relations.

(Suggestion: First try to show that if F is the free group on generators $x_1, \dots, x_n$, then the commutator subgroup of F cannot be generated as a normal subgroup of F by fewer than $n(n-1)/2$ elements. Then try to get the above statement from this.)

I.12:7. Lang asserts in the first Example on p.69 that in the group presented by three generators x, y, z and the three relations

(1) $\qquad [x,y] = y, \qquad [y,z] = z, \qquad [z,x] = x,$

one has $x = y = z = e$. Clearly, this is equivalent to saying that in any group with three elements x, y, z satisfying the above relations, $x = y = z = e$ holds.

Below, we shall prove this assertion, obtaining in the process some results about groups satisfying more general systems of relations. At one point we will call on a number-theoretic fact that uses an exercise from §II.1, but as we will note at the end, the case of this fact needed for the group described above can be checked by arithmetic. The exercise after this one outlines an alternative proof of Lang's assertion, with more work left to the reader.

(a) Verify that elements x and y of an arbitrary group satisfy

(2) $\qquad [x,y] = y$

if and only if they satisfy

(3) $\qquad xy = y^2x,$

equivalently, if and only if they satisfy

(4) $\qquad xyx^{-1} = y^2.$

(b) Suppose k is a positive integer, and G is a group containing elements $x_1, \dots, x_k$ satisfying the cyclic family of relations

(5) $\qquad [x_1,x_2] = x_2, \quad [x_2,x_3] = x_3, \quad \dots, \quad [x_{k-1},x_k] = x_k, \quad [x_k,x_1] = x_1.$

Show that if one of the x_i is of finite order, then all of them are of finite order. (Idea: If $x_i^m = e$, then conjugating x_{i+1} by x_i m times must leave it fixed; but m applications of (4) gives a different answer.)

(c) Assuming $x_1, \dots, x_k$ in (b) all have finite order, let $n > 1$ be a common exponent for these elements. By the same method used in (b), show that 2^n-1 is also a common exponent for these elements. But exercise II.1:2 shows that 2^n-1 is not divisible by n. Deduce that these elements have a common exponent smaller than n, and argue from this that they are all equal to e.

We can push this further ...

(d) Deduce that if a group contains elements satisfying (5), and a relation $x_i^a x_{i+1}^b = e$ holds with the integers a and b not both zero, then all x_j equal e. (Suggestion: Conjugate the given relation by x_i and compare the result with the original.)

... and still further ...

(e) Deduce that if a group contains elements satisfying (5), and a relation $x_i^a x_{i+1}^b x_{i+2}^c = e$ holds with a and b both positive, then all x_j equal e. (Suggestion: Consider two ways of transforming the given relation. On the one hand, apply (3), with x_{i+2} in the role of y, b times, to push the third term to the left past the second. On the other hand, apply (3), with x_{i+1} in the role of y, a times, to push the second term to the left past the first. Then use the general fact that in a group, $uv = e \Rightarrow vu = e$, to rearrange the result of the latter calculation so that the unchanged term x_i^a appears on the far left; equate this with the result of the former calculation, cancel x_i^a, and apply (d).)

We now come to the messy part: Showing that when $k = 3$, i.e., when (5) has the form (1), we automatically have a relation of the sort assumed in (e).

(f) Assuming (1), multiply the relation (3) on the left by z^4. Show with the help of (3), adjusted by cyclic permutations of x, y and z, that in the left-hand side of the resulting equation, you can convert a factor z^2x to x^4z^2, then a factor z^2y to yz, then a factor x^2y to y^4x^2. Likewise show that on the right you can convert a factor z^4y^2 to y^2z, and then a factor zx to x^2z. Equating the resulting expressions, deduce that $z^2x^2y^2 = e$, and complete the proof of Lang's assertion.

Remarks: When the argument of (b)-(e) is applied to the above relation $x^2y^2z^2 = e$, we find that the m of the case of (c) that arises is 6, and the only cases of II.1:2 that are needed are $\gcd(6, 2^6-1) = 3$, and $\gcd(3, 2^3-1) = 1$, which, of course, are verifiable without appealing to II.1:2.

For $k > 3$ the group presented by (5) can be shown to be nontrivial. By (c), this gives a nontrivial finitely presented group having no nontrivial finite homomorphic images.

I.12:8. This is an alternative proof of the assertion of Lang's proved in the preceding exercise. The part which we leave to your ingenuity is

*****(a) Suppose a group G contains elements x and y satisfying $[x,y] = y$. Show that $\langle x,y\rangle$ is solvable.

For the remainder of the proof, suppose G is generated by three elements x, y, z satisfying $[x,y] = y$, $[y,z] = z$, and $[z,x] = x$.

(b) Show from these relations that $G = [G,G]$. (Group-theorists call a group satisfying this condition *perfect*.)

(c) Also show from the given relations that the subgroup $\langle x,y\rangle$ contains the elements zxz^{-1}, z^2xz^{-2}, and z^2yz^{-1}.

(d) Verify that the product

$$(z^2yz^{-1})^{-1}(z^2xz^{-2})(z^2yz^{-1})(zxz^{-1})^{-1}(z^2yz^{-1})^{-1}$$

simplifies formally (without the use of the relations assumed) to $z(y^{-1}xyx^{-1}y^{-1})z^{-2} = z(y^{-1}[x,y])z^{-2}$. Conclude with the help of one of the given relations that $z \in \langle x,y\rangle$, and deduce that $\langle x,y\rangle = G$. Thus, by the result of (a) and this result, G is both perfect and solvable. However –

(e) Show that the only group that is both perfect and solvable is the trivial group.

II.1:1. (a) State and prove an analog of Goursat's Lemma (I:L5, p.75) for *rings*.

(b) Deduce a description of all subrings of the ring $\mathbf{Z} \times \mathbf{Z}$. (State your description in a form that is as simple and concrete as possible for this specific case. Recall that by ''ring'' we mean ''ring with 1'', and by ''subring'' we mean ''subring containing the same 1''.)

II.1:2. Show that if $n > 1$, then n is not a divisor of 2^n-1. (Hint: If n is odd, let p be the smallest prime dividing n, and show that the order of the image of 2 in the multiplicative group of $\mathbf{Z}/p\mathbf{Z}$ is not a divisor of n. What does this imply about divisibility properties of 2^n-1?)

Note on II:L12 (p.116): When Lang says in part (c) ''Let μ be the **Möbius function** such that ...'', he means ''Let μ be the **Möbius function**, i.e., the function defined by the equations ...''. (I do not recommend using ''such that'' in this way in your writing!)

The function μ so defined is a member of R, i.e., a map from P to K. In its definition, ''$(-1)^r$'' should thus be understood as $\pm$ the identity element $1 \in K$.

Lang asserts in part (c) that the result proved there gives the *Möbius inversion formula*. That formula is given as exercise V:L21 (p.254), which can be done right now, as an application of II:L12(c). (Lang has it in Chapter V because he wants to apply it in V:L22 to finite fields.) In doing it, you may assume that the abelian group A is the additive group of a ring. (In fact, any abelian group can be embedded in the additive group of a ring.)

Note on II:L13-19 (p.116), on Dedekind rings (defined on p.88): In the last line of the page, Lang points to VII:L7 for a continuation. There is also related material in III:L11-13.

II.2:1. Let R be a commutative ring. By a *submonoid* of R we shall understand a subset containing 1 and closed under multiplication. (Another term sometimes used is ''multiplicative subset''.) We use the word ''complement'' below in its set-theoretic sense.

(a) Show that the following five statements are *equivalent*.

(i) If I is an ideal of R, and S is maximal among submonoids of R disjoint from I, then the

complement of S in R is an ideal.

(ii) If S is a submonoid of R, and I is maximal among ideals of R disjoint from S, then the complement of I in R is a submonoid.

(iii) If I and S are an ideal and a submonoid of R, maximal for the property of being disjoint, then $I \cup S = R$. (Here ''maximal'' means among ordered pairs consisting of a disjoint ideal and submonoid of R, with respect to the partial ordering on such pairs that makes $(I_1, S_1) \leq (I_2, S_2)$ if and only if $I_1 \subseteq I_2$ and $S_1 \subseteq S_2$.)

(iv) If I and S are an ideal and a submonoid of R respectively, and are disjoint, then there exists an ideal $I' \supseteq I$ whose complement is a submonoid $S' \supseteq S$.

(v) If I and S are an ideal and a submonoid of R respectively, and are disjoint, then for every $r \in R$, either the ideal generated by I and r is disjoint from S, or the submonoid generated by S and r is disjoint from I.

(b) Prove (v) above, and hence all of the statements (i)-(v).
(Statement (ii) is II:L1, incidentally.)

(c) Show that an ideal of R is *prime* if and only if its complement is a submonoid of R.

Statement (iv) above, interpreted in the light of (c), is a very useful tool for constructing prime ideals. Here is an important application:

(d) Show that if I is any ideal of R, then the intersection of all prime ideals containing I is $\{r \in R \mid (\exists\, n \geq 0)\ r^n \in I\}$. (Hint: For every r not in the above set, you want to find an appropriate prime ideal that does not contain it.)

II.2:2. [> II.2:1]. Suppose R is a commutative ring, and $X \subseteq R$ a set having nonempty intersection with every nonzero prime ideal of R. Let $\langle X \rangle$ denote the multiplicative submonoid of R generated by X. Show that for every nonzero element $r \in R$, the set $\langle X \rangle$ contains some multiple of r.

II.2:3. [> II.2:1]. We know that an integral domain can be described as a commutative ring R in which $\{0\}$ is a prime ideal, and that in this situation, $\{0\}$ is the *least* prime ideal of R. If R is a commutative ring with zero-divisors, we can ask whether it still has a least prime ideal, and if not, whether it has *minimal* prime ideals (i.e., prime ideals not containing any smaller prime ideals).

(a) Show by example that a commutative ring need not have a least prime ideal.

(b) Prove that a nonzero commutative ring always has a minimal prime ideal. (Hint: Use I.2:1 to turn the question into one about the existence of a *maximal* set of some sort, and use Zorn's Lemma to prove the necessary existence statement.)

(c) Show that a prime ideal $P \subseteq R$ is minimal (among prime ideals) if and only if for every $a \in P$ there exists $b \in R - P$ and an integer $n > 0$ such that $a^n b = 0$.

II.3:1. (a) = IV:L8 (p.214). (Though Lang puts this in Chapter IV, it only requires the concepts of §II.3. The assumption that A be entire is not needed in this part, but will be needed in part (c) below.)

(b) Denoting the automorphism of $A[X]$ that carries X to $aX + b$ by $g(a, b)$, determine the law of composition of these automorphisms, and show that they form a group, which is a semidirect product of the additive group of A and the multiplicative group of units of A.

(c) = IV:L9 (p.214), but adding after ''automorphism of $A[X]$'' the words ''inducing the identity on A''. (If you have seen §III.1, this means we are looking at automorphisms as an A-*algebra*.)

(d) Show that if A is a field, then every endomorphism of $A[X]$ *as a ring* carries A into itself, and that every *automorphism* of $A[X]$ as a ring carries A *onto* itself.

(e) Show that if A is a field, the group of all automorphisms of $A[X]$ as a ring is a semidirect product of the group of automorphisms of the field A, and the group of A-algebra automorphisms of $A[X]$,

described in (a-c) above. What is the action of the former group on the latter?

II.4:1. Prove the assertion in the second paragraph of p.90 of Lang, that every field of characteristic 0 contains a subfield isomorphic to $\mathbf{Q}$. (You may use the results of §II.4, and the fact that the field of fractions of $\mathbf{Z}$ is $\mathbf{Q}$.)

II.4:2. (a) Show that a commutative ring is local if and only if its noninvertible elements form an ideal.

(b) = II:L3 (p.115).

II.4:3. [> III.2:1-2]. This exercise shows how to define the concept of not necessarily commutative local ring. Let A be a ring.

(a) Show that the following conditions are equivalent:

(i_{left}) A has a unique maximal left ideal.

(i_{right}) A has a unique maximal right ideal.

(ii) The set of noninvertible elements of A forms a 2-sided ideal.

(Suggestion: Translate (i_{left}) into a statement about left annihilators $\{a \in A \mid ax = 0\}$ of elements x of simple left A-modules.)

A ring satisfying the above equivalent conditions is said to be *local*. There are some more equivalent conditions, which for completeness I give below.

(b) Show that the conditions of part (a) are also equivalent to each of the following:

(iii_{left}) The set of non-left-invertible elements of A is closed under addition.

(iii_{right}) The set of non-right-invertible elements of A is closed under addition.

(iv) The set of noninvertible elements of A is closed under addition.

(Hint: Show that a ring satisfying (iv) cannot contain any idempotent elements, i.e., elements e satisfying $e^2 = e$, other than 0 and 1, but that a ring with right-invertible but not left-invertible elements does have such idempotents.)

In an earlier version of this Companion, I included another condition which I thought was equivalent to the above, but turned out not to be:

(c) Show that conditions (i)-(iv) above are *not* equivalent to:

(v) A has a unique maximal 2-sided ideal $\mathbf{m}$, and $A/\mathbf{m}$ is a division ring.

(If you haven't done all of parts (a) and (b), you can still do this part by obtaining a ring for which the truth-value of (v) is different from *at least one* of the other conditions listed.)

II.4:4. In the last paragraph of p.110, Lang gives three equations which say that the operation of ''localizing ideals'' respects the operations of adding, multiplying, and intersecting ideals. More generally, given any homomorphism of commutative rings, $f : A \to B$, let us, as Lang does there, write $J(A)$ for the set of all ideals of A, and let us define $f_* : J(A) \to J(B)$ to take each ideal $\mathbf{a} \in J(A)$ to the ideal $B f(\mathbf{a}) \in J(B)$, and $f^* : J(B) \to J(A)$ to take $\mathbf{a} \in J(B)$ to $f^{-1}(\mathbf{a}) \in J(A)$.

(a) Of the three equations Lang proves for the operation ''$S^{-1}(-)$'', determine which hold for the operation f_* for every homomorphism f of commutative rings. Prove the equations that hold, and give counterexamples for those that don't.

(b) Do the same for the operation f^*.

(c) Show that the operation $S^{-1}(-)$ treated by Lang is in fact f_* for f the canonical map $A \to S^{-1}A$. Assuming the results that you proved in (a) above, is the one verification given by Lang at the end of that paragraph sufficient to establish the three equations he shows?

(d) Again taking for f the canonical map $A \to S^{-1}A$, does the map f^* satisfy some or all of the same three equations?

(e) (Open-ended) What properties can you prove for the composite maps $f^* f_*$ and $f_* f^*$, either in general, or for the specific cases where f is the canonical map $A \to S^{-1}A$?

II.5:1. Let k be any field, and $k[x]$ the ring of polynomials over k in one indeterminate x. Let $k[x^2, x^3]$ denote the subring of $k[x]$ consisting of those polynomials $\Sigma\, a_i x^i$ such that $a_1 = 0$, i.e., polynomials with zero linear term. (The symbol $k[x^2, x^3]$ means that this subring is generated over k by the two elements x^2 and x^3, but I do not ask you to verify this fact.)

(a) Establish necessary and sufficient conditions on nonnegative integers $i, j \neq 1$ for x^i to divide x^j in $k[x^2, x^3]$. Show by example that this is *not* equivalent to the condition that x^i divide x^j in $k[x]$.

(b) Show that the elements x^2 and x^3 have a greatest common divisor in $k[x^2, x^3]$, but not a least common multiple. (You may take it for granted that any divisor of an element x^j is, up to multiplication by a unit, of the form x^i. Remember that ''least'' and ''greatest'' refer here to divisibility in this ring, not in $k[x]$.)

(c) Suppose a and b are any nonzero elements of any integral domain R. Show that if u and v are elements of R such that $ab = uv$, then u is a common divisor of a and b if and only if v is a common multiple of a and b. Show moreover that if u' and v' are two other elements such that $ab = u'v'$, then u is a multiple of u' if and only if v' is a multiple of v. Show, finally, that a *least common multiple* of a and b, if it exists, will be a divisor of ab, and that the same is true of a *greatest common divisor* of a and b if it exists. (In proving these statements, note that you may *not* assume that elements of the ring R have unique factorization, or even admit any factorization into irreducibles.)

(d) The results of (c) appear to imply that two nonzero elements a and b of an integral domain should have a *least* common multiple if and only if they have a *greatest* common divisor. But this contradicts the example of (b)! Find the fallacy; and verify that one direction of this ''if and only if'' *is* valid.

(e) Find two elements of $k[x^2, x^3]$ that have no greatest common divisor. (Suggestion: two common multiples of x^2 and x^3.)

II.5:2. Let R be a commutative integral domain. By a *prime element* of R we will mean a nonzero element p such that pR is a prime ideal. Show that the following conditions are equivalent:

(i) R is a unique factorization domain.

(ii) Every irreducible element of R is prime, and there is no infinite strictly increasing chain of principal ideals, $Ra_1 \subsetneqq Ra_2 \subsetneqq \dots$.

(iii) Every nonzero nonunit in R is a product of prime elements.

II.5:3. [> II.2:1, II.5:2] In the context of II.5:2, show that conditions (i)-(iii) of that exercise are also equivalent to each of:

(iv) Every nonzero prime ideal of R contains a nonzero principal prime ideal.

(v) R has no infinite strictly increasing chain of principal ideals, $Ra_1 \subsetneqq Ra_2 \subsetneqq \dots$, and satisfies the *factorization refinement property:* For any relation $ab = cd$ holding among nonzero elements of R, there exist elements $w, x, y, z \in R$ such that $a = wx$, $b = yz$, $c = wz$, $d = xy$.

Suggestions for proving that (iv) implies one of the other conditions: Use II.2:2 with X the set of nonzero prime elements of R, or II.2:1 with S the monoid generated by these elements. You may also want to review how Lang proves that a PID (in his language, a ''principal ring'') is a UFD (a ''factorial ring''), though he doesn't do all parts of that proof in the smoothest possible way.

***II.5:4.** Find an example of a commutative ring R with elements x and y such that $xR = yR$, but such that there does not exist a unit $u \in R$ such that $y = ux$.

II.5:5. Let A be an integral domain and K its field of fractions.

(a) Show that if A is a PID, then every subring R of K containing A is a localization $S^{-1}A$ for

some multiplicative set S.

(b) Show that if A is a UFD such that every subring R of K containing A is a localization $S^{-1}A$ for some multiplicative set S, then A is a PID.

Clarification of II:L5 (p.115): Change ''with primes p'' to ''where p is a prime'', and ''units u'' to ''where u is a unit''.

II.5:6 [> II:L4, II:L10]. Let R be the ring of trigonometric polynomials, defined in the last paragraph before the exercises on p.114, and studied in II.5:10. Let $R[i]$ denote the ring of complex-valued functions on the real line of the form $u + v\,i$ with $u, v \in R$.

(a) Show that $R[i]$ is generated over $\mathbf{C}$ by the function e^{ix} and its multiplicative inverse.

(b) Show that $R[i]$ is a localization of a polynomial ring, and deduce that it is a principal ideal domain.

For the remainder of this exercise, let $I_0 \subseteq R$ denote the set of trigonometric polynomials $f \in R$ that satisfy $f(0) = 0$.

(c) Show that I_0 is a non-principal ideal of R.

Thus, R is not a principal ideal domain. However, the remaining parts will show that it has properties very close to those of a principal ideal domain.

(d) Show that for every ideal $J \subseteq R$, there exists a unique ideal $J' \subseteq R[i]$ closed under complex conjugation and satisfying $J' \cap R = J$. (Remark: This will of course remain true if we delete the ''closed under conjugation'' condition; but then uniqueness can fail. You might look for an example of this.)

(e) If we apply part (d) to the ideal I_0, the corresponding ideal of $R[i]$ must be principal. Find a generator (or better: all possible generators) of this ideal. Why do these elements fail to give generators of I_0 over R, contradicting (c)?

(f) Let J be a nonzero ideal of R, J' the ideal of $R[i]$ described in (d), and $f \in R[i]$ a generator of J'. Let n be the number of zeroes of f on the interval $[0, 2\pi)$, counting multiplicity. Show that the ideal J is principal if and only if n is even.

(g) Deduce that for every nonzero ideal $J \subseteq R$, exactly one of the ideals J and $I_0 J$ is principal.

Remark: From (g), one can deduce that R is a *Dedekind domain*, a concept briefly mentioned by Lang under the name ''Dedekind ring'' on p.88, and studied further in Exercises II:L11-19 and other exercises in later chapters. The idea that principal ideals of R are to all ideals as even integers are to all integers, suggested by (g), can be made precise by deducing from (g) that the *ideal class group* of R (also defined on p.88) is isomorphic to Z_2.

Historically, Dedekind domains arose in number theory, and the examples that specialists are most familiar with come from that subject. I think they would consider this one a somewhat exotic example.

Notes on exercises in Lang: II:L9 belongs under Chapter IV. It is essentially worked in this Companion, immediately following Theorem IV.1.c2.

Hint to II:L10(c): If $a + b\sqrt{(-D)}$ is a unit of R, consider the product of this element with its complex conjugate.

III.1:1. [> §II.4]. As noted in the remarks on this section, we will call a morphism $f\colon A \to B$ in a category an ''epimorphism'' if for any object C and any morphisms $g, g'\colon B \to C$, one has $gf = g'f \Rightarrow g = g'$, rather than using the meaning Lang gives this term.

(a) Show that in the category of left modules over a ring R, the epimorphisms are precisely the surjective homomorphisms. (Hint: If f is not surjective, take for g and g' a zero map and a nonzero map into an appropriate object.)

(b) Show, on the other hand, that if R is a commutative ring and S a multiplicative system in R, then

the natural map $R \to S^{-1}R$ is an epimorphism in the category of commutative rings, but that such maps are not in general surjective.

III.1:2. In a ring R, the *centralizer* of a subset X is defined (as in a group or semigroup) to be $\{r \in R \mid (\forall\, x \in X),\ rx = xr\}$.

(a) Show that the centralizer of any subset of a ring is a subring.

(b) Suppose R is a ring, M an R-module, $\mathrm{End}_{\mathbf{Z}}(M)$ the ring of *abelian group endomorphisms* of M, and $\mathrm{End}_R(M)$ the ring of *R-module endomorphisms* of M. Recall that the R-module structure of M is given by a homomorphism $\varphi\colon R \to \mathrm{End}_{\mathbf{Z}}(M)$. Show that $\mathrm{End}_R(M)$ is the centralizer of $\varphi(R)$ in $\mathrm{End}_{\mathbf{Z}}(M)$.

(If you have read as far as my comment to p.124 of Lang, to the effect that many noncommutative ring-theorists follow the convention that endomorphisms of right modules are written to the left of module elements and vice versa, note that we are *not* following this convention in this exercise – since we are here looking at ring-element actions and module-endomorphisms as members of a common ring $\mathrm{End}_{\mathbf{Z}}(M)$!)

The remaining point is an example of a recurring mathematical pattern, worth noting.

(c) Let R be a ring, and for every subset $X \subseteq R$, let us abbreviate the centralizer of X in R by X^*. Show that for every X, we have $X^{**} \supseteq X$ and $X^{***} = X^*$.

(When R is the ring of endomorphisms of an abelian group, and X is a subring thereof, so that we are in the situation of the preceding parts, then X^* is called the *commutant* of X, and X^{**} the *bicommutant*.)

III.1:3. In the *Sketch of Proof* of Lemma III.1.c1 in this Companion, the words ''It is easy to verify'' occur three times. Give these three verifications.

Note on III:L10 (p.168, based on §III.1) [> III:L9, p.167]. Suggested plan of attack: Given a multiplicative set S, determine what elements of M go to zero under the natural map $M \to S^{-1}M$. Hence, given a prime ideal $\mathbf{p}$, determine what elements go to zero under $M \to M_{\mathbf{p}}$. Then show that for every nonzero $x \in M$, there is a maximal ideal $\mathbf{p}$ which does not make x go to zero.

III.1:4. Let A be a ring. In the next-to-last paragraph of p.118, Lang notes that if A is an integral domain and M an A-module, then the set of elements $x \in M$ whose annihilator, $\mathrm{ann}_A\, x = \{a \in A \mid rx = 0\}$, is nonzero, forms a submodule M_{tor} of M. However, the arguments showing this fail if either the assumption that A is commutative or the assumption that it has no zero-divisors is deleted.

However, for general A, there may still be one or more ways of describing a submodule of each A-module consisting of all elements whose annihilators have certain properties.

So – determine necessary and sufficient conditions on a set X of left ideals of A for it to be true that for every left A-module M, the set of elements of M whose annihilators belong to X forms a submodule M_X of M.

III.2:1. A module over a ring R is called *simple* if it is nonzero, but has no nonzero proper submodules. Show that if M is a simple R-module, then $\mathrm{End}_R(M)$, the ring of endomorphisms of M as an R-module, is a division ring.

III.2:2. For any element x of a left R-module M, one defines the *annihilator* of x by

$$\mathrm{ann}_R\, x \;=\; \{r \in R \mid rx = 0\}.$$

(a) Show that in this situation, $\mathrm{ann}_R\, x$ is a left ideal of R, and construct a module isomorphism between the submodule $Rx \subseteq M$ generated by x, and the left R-module $R/\mathrm{ann}_R\, x$. (Thus, I am asking you to verify some things I asserted without proof in a comment to p.118 of Lang in this Companion.)

A left module over a ring R is called *cyclic* if it is generated by a single element. The above result says that the structure of a cyclic R-module is determined by the annihilator of any generator.

(b) Let Rx and Ry be cyclic R-modules, with $\mathrm{ann}_R\, x = I$ and $\mathrm{ann}_R\, y = J$. Clearly, a homomorphism $Rx \to Ry$ takes x to an element of the form ay, and is therefore determined by the element a. Characterize, in terms of the left ideals I and J, those elements $a \in R$ such that there exists a homomorphism $Rx \to Ry$ taking x to ay, and determine when the homomorphisms corresponding to two such elements a and b are the same. Thus, get a description of the abelian group $\mathrm{Hom}(Rx, Ry)$.

(c) Obtain from the above result a description of the endomorphism ring $\mathrm{End}_R(Rx)$ of a cyclic R-module Rx in terms of $I = \mathrm{ann}_R\, x$. In this case, you must, of course, determine its multiplication as well as its additive group structure.

(d) [cf. III.2:1] Show that the *simple* R-modules are precisely the cyclic R-modules Rx such that $\mathrm{ann}_R\, x$ is a maximal left ideal of R.

(e) In the case where R is commutative, state the forms that the results of (b) and (c) above take, and with the help of (d) above, describe the division rings that occur as in III.2:1.

III.2:3. [cf. III.2:1]. Part (a) below gives a class of examples showing that the converse to III.2:1 is not true. Parts (b) and (c) are more difficult questions that this result suggests.

(a) Show that if R is a commutative integral domain which is not a field, and K is its field of fractions, then the underlying R-module of K is not simple, but its ring of R-module endomorphisms is isomorphic to the field K.

*****(b) Find a *finitely generated* module M over a ring R such that $\mathrm{End}_R(M)$ is a division ring (possibly a field), but such that M is not simple.

To state the last part of this exercise, let us generalize the construction of part (a) as follows. Suppose $f\colon A \to B$ is any homomorphism of rings, and let us regard B as a left A-module by defining left multiplication by $a \in A$ to take $x \in B$ to $f(a)x$. Then we see that for every $b \in B$, the map $x \mapsto xb$ gives an endomorphism of B as left A-module. For convenience, let us write left A-module endomorphisms of B on the *right* of the elements they are applied to, and compose them accordingly. Then the above map is a ring homomorphism $B \to \mathrm{End}_A(B)$. This is clearly one-to-one, but let us ask, open-endedly,

*****(c) For what ring homomorphisms $f\colon A \to B$ is the homomorphism $B \to \mathrm{End}_A(B)$ described above an isomorphism?

III.3:1. Let R be any ring, and $0 \to L \to M \to N \to 0$ a short exact sequence of left R-modules. Show that the following five conditions are equivalent:

(i) For every left R-module A, the induced sequence $0 \to \mathrm{Hom}(A, L) \to \mathrm{Hom}(A, M) \to \mathrm{Hom}(A, N) \to 0$ is exact.

(i′) The induced map $\mathrm{Hom}(N, M) \to \mathrm{Hom}(N, N)$ carries some element of the former abelian group to the element id_N of the latter abelian group.

(ii) For every left R-module A, the induced sequence $0 \leftarrow \mathrm{Hom}(L, A) \leftarrow \mathrm{Hom}(M, A) \leftarrow \mathrm{Hom}(N, A) \leftarrow 0$ is exact.

(ii′) The induced map $\mathrm{Hom}(L, L) \leftarrow \mathrm{Hom}(M, L)$ carries some element of the latter abelian group to the element id_L of the former abelian group.

(iii) The original exact sequence of R-modules splits (i.e., satisfies the equivalent conditions of Proposition III.3.2, p.132. Cf. the discussion of splitting in the context of abelian groups in this Companion, after the notes on the proof of Lemma I.7.2.)

III.3:2. Let A be a ring.

(a) Show that if an A-module M is the direct sum of submodules $M_1, \ldots, M_n$, then its A-module structure can be extended to a structure of $(A \times \ldots \times A)$-module (direct product ring with n factors) in such a way that the element of this ring having a 1 in the ith position and zeros elsewhere acts as the projection

onto the summand M_i; and conversely, that every $(A\times\ldots\times A)$-module arises in this way from a direct sum decomposition.

(b) Show that if I is an infinite set, and we denote by A^I the direct product of an I-tuple of copies of A, then any A-module with a decomposition as a direct sum of an I-tuple of submodules can be given an A^I-module structure in a way analogous to the above, and that the same is true of any A-module with an expression as a direct product of an I-tuple of modules, but that there are also A^I-modules that do not arise in either of these ways.

III.4:1. Let R be a nonzero ring, M a left R-module, and X a subset of M, and let RX denote the submodule of M generated by X. Show that X is contained in a *basis* of M if and only if X is linearly independent and the factor-module M/RX is free.

***III.4:2.** Let R be the ring of all continuous real-valued functions on the interval $[0,1]$, and consider the ideal $I = \{\, f\in R \mid (\exists\varepsilon>0)\ (\forall x\in[0,\varepsilon])\ \ f(x) = 0\}$. Show that as an R-module, I is projective.

Note on III:L5: In giving a proof that follows Lang's sketch, you need to explain his ''we may assume'' step; i.e., to say how, given v_1, ..., v_m, and assuming his inductive hypothesis, you can modify v_1, ..., v_m to get a family with the property that he says may be assumed.

III.5:1. We noted in this Companion that if R is a ring, n a positive integer, and A the ring of $n\times n$ matrices over R, and we regard the set of column vectors of height n over R as a left A-module P, then A is isomorphic as a left A-module to the direct sum of n copies of P.

Now suppose that R is a field, or more generally, a division ring, let n and A be as above, and let M be any left A-module. Let us write e_{ij} for the usual matrix units in A.

(a) Show that $e_{11}M$, i.e., $\{e_{11}x \mid x\in M\}$, may be regarded as a left R-vector-space.

(b) Let X be any basis of the above vector space. Show that for each $x\in X$, the left A-module Ax is isomorphic to P, and that M is the direct sum over X of these modules.

Thus, every A-module is projective, and is a direct sum of copies of P. (A more general result of this sort is developed in Exercise XIII.3:1.)

III.5:2. (a) Show that for any positive integer n and infinite field k, the following statements are equivalent.

(i) An affirmative answer to Open Question III.5.c4(a).

(ii) An affirmative answer to the same question with the assumption that V is n-dimensional replaced by the statement that it has dimension at least n, and ''bases'' replaced by ''linearly independent subsets''.

(iii) As in (ii), but this time with n-dimensional replaced by $\leq n$-dimensional, and ''bases'' replaced by ''spanning subsets''.

(Remark: The assumption that k is infinite is convenient here. However, using the fact that any field k can be embedded in an infinite field, one can show that if the above results are true for all infinite fields k, they are in fact true for all fields k.)

(b) Show that the above equivalent statements imply that given a set S, and n^2 n-element subsets S_{11}, $S_{12}, \ldots, S_{nn}$, one can choose an element s_{ij} from each S_{ij} in such a way that for each i the n elements s_{ij} $(j=1,\ldots,n)$ are distinct, and for each j the n elements S_{ij} $(i=1,\ldots,n)$ are distinct.

The assertion of (b) was a longstanding open question, called the Dinitz Conjecture, till it was proved by Fred Galvin (*The list chromatic index of a bipartite multigraph*, J. Combin. Theory, Ser. B, **63** (1995) 153-158). The vector-space questions are still open, to the best of my knowledge.

Correction to III:L6 (p.166): At the start of the next-to-last line, before ''a $\mathbf{Z}$-basis'' add ''a subgroup of finite index with''. (The statement in Lang is false. The above modified statement is almost trivial to prove, once one sees the method. I don't know any version of the statement that is correct but nontrivial

to prove.)

Note also that in the next-to-last line, the subscript on $\{y_s\}$, which appears as *wes* through the first Springer printing, and as $w \in s$ in the 4th Springer printing, should be $w \in S$.

III.6:1. Let A be a commutative ring, B a commutative A-algebra, and M a B-module. Recall from Lemma III.1.c3 in this Companion (notes to Lang's p.121) that the B-module M may be regarded as an A-module by letting each $a \in A$ act on elements of M via the image of a in B. Let us denote by $M^\vee$ the A-module $\mathrm{Hom}_A(M, A)$.

(a) Show that $M^\vee$ can be made a B-module by defining, for each $b \in B$ and $f \in M^\vee$, the element $bf \in M^\vee$ via the equation $(bf)(x) = f(bx)$.

(b) Suppose A is a field, and b any element of B. Show that multiplication by b is one-to-one on $M^\vee$ if and only if it is surjective on M, and that it is surjective on $M^\vee$ if and only if it is one-to-one on M.

(c) Still letting A be a field, let $B = A[X]$, the polynomial ring over A, and let $M = B$ regarded as a module over itself. We see from part (b) that $M^\vee$ will contain nonzero elements annihilated by X and by $X+1$. Find such elements.

(Remarks: If we had not assumed B commutative, we could still have gotten a result like (a), except that a *left* B-module structure on M would have yielded a *right* B-module structure on $M^\vee$. More generally, for any A-module N and left B-module M we get a right B-module structure on $\mathrm{Hom}_A(M, N)$. If, in addition, N is a left B-module, then this induces a left B-module structure on the same set, such that these two structures together make it a (B, B)-bimodule. Within that bimodule $\mathrm{Hom}_A(M, N)$, one can characterize $\mathrm{Hom}_B(M, N)$ as a ''centralizer'', $\{f \mid (\forall\, b \in B)\ bf = fb\}$.)

III.6:2. Show that if P is a finitely generated projective left module over a ring R, then the right R-module $P^\vee = \mathrm{Hom}(P, R)$ is also finitely generated and projective, and the natural map $P \to P^{\vee\vee}$ is an isomorphism.

Note on exercises for §III.7: Most of the exercises given earlier in this Companion for §I.8 can be used as exercises for this section, with ''abelian group'' everywhere replaced by ''A-module'', where A is a principal ideal domain. For those exercises that ask for counterexamples, some additional assumptions are needed (since when A is a field, the statements to which counterexamples are being asked for tend to be true). In these cases, it might be made part of the exercises to find conditions on a principal ideal domain under which such counterexamples will exist.

III.6:3. Let k be a field, let $f: U \to V$ be a linear map of K-vector-spaces, and let $f^\vee: V^\vee \to U^\vee$ be the induced homomorphism of dual spaces.

(a) Show that $f^\vee$ is one-to-one if and only if f is onto, and that $f^\vee$ is onto if and only if f is one-to-one. (This will be easy, using an appropriate result of this section.)

(b) Regarding each of the if-and-only-if assertions of part (b) as two statements, determine which of the resulting four statements remain true if ''field K'' and ''vector space'' are replaced by ''ring A'' and ''A-module''.

(Insofar as you find counterexamples, you might see whether you can get them to satisfy to various additional conditions, such as using modules that are finitely generated, and/or free, and/or with base-ring A the integers.)

III.7:1. Let F be a free module of finite rank over a principal ideal domain R, and $D \subseteq E \subseteq F$ submodules. Must there exist a basis $\mathfrak{B}$ of F such that both D and E are spanned by multiples of members of F? (Cf. Theorem III.7.8.) Give a proof or a counterexample.

III.7:2. Theorem III.7.9 in our text (as corrected in the first Springer printing) has the hypothesis that the

elementary matrices in R generate $GL_n(R)$. This leads to the question: For which R is this true?

Show that this condition holds for $R = \mathbf{Z}$.

Suggestion on how to start: Given an element of $GL_n(R)$, show that by row operations one can reduce the sum of the absolute values of the entries in the first column to 1, and finally bring that column to a form with 1 in the first position and zeroes below. Show that the submatrix below and to the right of that ''1'' is then invertible.

We will generalize this result in Exercise IV.1:3.

Note on III:L14 (p.169, based on §III.9): Lang's ''Hint'' should be considered to apply to the exercise as a whole, not to part (c) in particular. In fact, it appears to me that part (c) is most easily done by using parts (a) and (b) to get one case, and a non-snake argument in the remaining two easier cases.

Note on III:L16 (p.169, based on §III.10): I would mark this ''¤''; i.e., it is essentially trivial once one makes an easy observation (at least as it is stated in all printings through the 4th Springer printing: ''Prove that the inverse limit of a system of simple groups in which the homomorphisms are surjective is either the trivial group, or a simple group''). Perhaps what Lang intends is, rather, the following. Note that here, no surjectivity assumption is made.

III.10:1. Prove that the *direct* limit of a system of simple groups is either the trivial group, or a simple group.

On the other hand, if we consider inverse limits, but delete the surjectivity assumption, we can pose

***III.10:2.** Show by example that an inverse limit of simple groups need *not* be simple.

IV.1:1. = the handout *A principal ideal domain that is not Euclidean*. (When I assign this, I generally use parts (1)-(7) as two thirds of a weekly homework assignment.)

***IV.1:2.** Suppose A is a commutative ring, and $f \in A[X]$ is a zero-divisor. Show that there exists an element $a \in A - \{0\}$ such that $af = 0$.

IV.1:3. Do Exercise III.7:2 for the more general case where R is a Euclidean domain (defined in these notes, re p.175).

IV.3:1. Part (a) below generalizes the Eisenstein Irreducibility Criterion (Theorem IV.3.1, p.183). As in the statement of that result, let A be a unique factorization domain, K its field of fractions, p a prime element of A, and $f(X) = a_n X^n + \ldots + a_0 \in A[X]$.

(a) Suppose $1 \leq j < m \leq n$ are such that $a_i \equiv 0 \pmod{p}$ for all $i < m$, not all coefficients of f are divisible by p, and $a_0 \not\equiv 0 \pmod{p^{j+1}}$. Show that when f is factored into irreducibles in $K[X]$, there will be a family of at most j of these factors whose product has degree at least m. (Hint: Look at those factors of f whose constant term is divisible by p. Compare the factorization of f in $A[X]$ and its image in $(A/(p))[X]$.)

(b) Obtain Eisenstein's Criterion as a special case.

(c) Deduce, as another case, that if in the statement of Eisenstein's Criterion, the condition $a_i \equiv 0 \pmod{p}$ is only assumed for $i < n-1$, then f is either irreducible, or the product of a linear factor and an irreducible factor of degree $n-1$.

IV.4:1. If R is a ring, I an ideal of R, and S a subring of R, it is not hard to verify that the set $S + I = \{s + i \mid s \in S, i \in I\}$ forms a subring of R. I will not ask you to prove this straightforward fact; but let us see how this construction can be used to get examples of non-Noetherian rings.

(a) Show that the subring $\mathbf{Z} + x\mathbf{Q}[x]$ of $\mathbf{Q}[x]$ (consisting of all polynomials with rational coefficients, and constant term an integer) is not Noetherian.

(b) Let k be a field. Show that the subring $k + xk[x, y]$ of $k[x, y]$ (consisting of all elements in

which the coefficient of each monomial y^i ($i > 0$) is zero) is not Noetherian.

*****(c) Show that the ring constructed in (a) above has the property that every *finitely generated* ideal is principal, but is not a UFD.

*****(d) Show that if k is a field and n a positive integer, then $k + x^n k[x]$ is Noetherian, and in fact has the property that every ideal can be generated by $\leq n$ elements, but contains an ideal which cannot be generated by $< n$ elements.

IV.6:1. = the handout on *Solutions in radicals*. (Although this exercise seems naturally to go with §VI.7, it can be assigned any time after §IV.6, provided the terms ''primitive cube roots and fourth roots of unity'' are explained and their elementary arithmetic properties noted. I like to assign it early, and then discuss it again after we read §VI.7. As most of a week's assignment, I typically assign parts (1)-(4) and a choice of (5) or (6). Part (8) depends on material that is in Hungerford but not Lang, and so should not be assigned unless this is covered separately. Concerning part (2), the instructor should specify what results about generating sets for S_n, e.g., I:L38(a-c) and/or I.5:7, may be assumed.)

Note on IV:L12(c): The notation D_f for the discriminant is as defined on p.193; on p.204 the same element is written $D(f)$.

V.1:1. = the handout on *Lüroth's Theorem*. (As a week's assignment, I typically give parts (1)-(9). Part (13) is immediate from II.5:1 above, and so should be omitted if that exercise has been assigned.)

In the phrase ''simple transcendental extension'', defined on that handout, the word ''simple'' has its original sense ''one-fold'', in this case ''generated by a single element''. One does not call this a ''cyclic extension'' because in Chapter VI, that term will be used to mean an extension having cyclic Galois group.

¤**V.1:2.** Suppose E and F are subfields of a field K, such that every element of E is algebraic over F, and every element of F is algebraic over E. Must E and F be algebraic over $E \cap F$?

Notes on a couple of Lang's exercises in Chapter V. One direction of V:L17 (p.254, based on §V.6) is a result in the text, so the exercise is essentially to prove the converse. Exercise V:L21 is an immediate consequence of II:L12; see note on that exercise at end exercises for II.1. Hint to V:L28 (p.256, based on §V.2): Reduce to the case where the extension of odd degree has the form $k(\alpha)$, express this as $k[X]/(f(X))$, and thus translate the hypothesis into a statement about polynomials. Expect to use induction on the degree of f.

Note: Exercises VI.14:1 and VI.14:2 can actually be given at any point after §VI.1; cf. comment preceding those exercises.

VI.1:1. Let K be a finite Galois extension of a field k, with Galois group G. Since G is a group of permutations of K, we may regard k as a G-set. Let $\alpha, \beta \in K$.

(a) Show that the orbits $G\alpha = \{\lambda\alpha \mid \lambda \in G\}$ and $G\beta = \{\lambda\beta \mid \lambda \in G\}$ are isomorphic as G-sets by an isomorphism that carries α to β if and only if $k(\alpha) = k(\beta)$.

(b) Show that $G\alpha$ and $G\beta$ are isomorphic as G-sets by some isomorphism if and only if $k(\alpha) \cong k(\beta)$ as extensions of k.

(c) In the situations of (a) and (b), what interpretation does the cardinality $|G\alpha|$ have in terms of the extension $k(\alpha)$ of k?

(d) Find characterizations analogous to (a) and (b) of the situation where $G\beta$ is the image of $G\alpha$ under a homomorphism of G-sets, with and without the condition that the homomorphism carry α to β.

(e) Why have I not also asked the analog of question (d) with ''homomorphic image'' replaced by ''sub-G-set''?

The only part of §VI.2 that the next exercise assumes is Example 1 on p.269, so it can be given after

covering §VI.1, with a reference to that example. I typically give this exercise as a full week's homework assignment.

VI.2:1. = VI:L8 (p.322). However, in parts (a) and (b), assume that you are working over an arbitrary field k of characteristic $\neq 2$, rather than $\mathbf{Q}$. Also begin by justifying (briefly) the statement that the roots of f may be written $\pm\alpha$, $\pm\beta$.

If you have a printing before the 4th Springer printing, there are errata to part (b) for which you should see the pages of errata. (There are three errata-listings there. Which one applies depends on the printing you have.)

In part (c), each subextension other than the base field k itself will be quadratic over an extension just below it in the lattice, and hence can be obtained by adjoining to that extension the square root of an element thereof (p.269, Example 1). Hence, try to express each subextension as obtained from k via successive adjunctions of square roots.

In doing this exercise, you may take for granted that the group that Lang calls D_8 is a 2-Sylow subgroup of S_4, and can be pictured as the group of symmetries of the square by identifying the four elements on which S_4 acts with the vertices of the square. In (b), expect to use repeatedly the observation that an element lies in k if and only if it is fixed by all members of the Galois group. All the assertions in that part, including the parenthetical one, are to be proved, even though only one assertion is preceded by ''Show that''. In (c) you may take for granted that the subgroups of D_8 are as shown in the chart on p.271, for appropriate choices of generators σ and τ satisfying the relation displayed a few lines above the diagram. In the last part of (c), where you are asked for the field associated with each subgroup of the Galois group, you may find it useful to apply the result of (b) not only over $\mathbf{Q}$, but also over some of the fields you are trying to determine.

In displaying the diagram of subextensions, you will have to choose between the arrangement that parallels the diagram of subgroups (with *smaller* subextensions, corresponding to larger subgroups, at the top) and the arrangement natural for these subextensions, with *larger* subextensions at the top. To give the grader a uniform task, use the latter arrangement. For the same reason, take τ to be complex conjugation.

The word ''lattice'' in the last sentence of (c) means ''partially ordered set in which every pair of elements has a least upper bound and a greatest lower bound''; but you are not asked to prove anything about such upper and lower bounds.

VI.2:2. (converse to VI.2:1(a), and related results). Let k be a field not of characteristic 2.

(a) Suppose K is a Galois extension of k whose Galois group G is isomorphic to the group Lang writes D_8. Show that K is the splitting field over k of an irreducible polynomial $f(X)$ of the form $X^4 + aX^2 + b$. (Hint: Find a non-normal subgroup $H < D_8$ of order 2, let N be the normal subgroup it generates, and note that K^H is generated by the square root of an element $\gamma \in K^N$.)

(b) Show that the above conclusion is also true of any Galois extension K of k having degree 4. (Suggestion: Consider separately the cases of Galois group Z_4, and $Z_2 \times Z_2$.)

(c) Show that if an irreducible polynomial f of degree 4 has splitting field of degree not divisible by 3, then its Galois group is D_8, Z_4, or $Z_2 \times Z_2$. Must such an f have the form $X^4 + aX^2 + b$?

(d) Show that if f is an irreducible polynomial over any field whose Galois group is the 8-element ''quaternion group'' described by Lang on p.9, then f must have degree 8.

(The study of polynomials having this Galois group is more difficult than the case of D_8, and will not be pursued further here.)

Still another variant of the theme of this and the preceding exercise is given in VI.2:6.

VI.2:3. [cf. I.5:6]. Let f be a separable irreducible polynomial of degree $n > 1$ over a field k, let K be the splitting field of f, and let $\alpha_1, \ldots, \alpha_n$ be the roots of f in K. Show that the following conditions are equivalent:

(i) $f(X)/(X-\alpha_1)$ is irreducible in $k(\alpha_1)[X]$.

(ii) $G(K/k)$ is 2-transitive on $\{\alpha_1, \ldots, \alpha_n\}$ (as defined in Exercise I.5:6).

VI.2:4. Let k be a field, f a separable irreducible quadratic polynomial over k, and g a separable irreducible cubic polynomial over k. Let us regard the Galois group G of fg as a group of permutations of the five roots of this polynomial. Since it permutes the set of roots of f and g separately, it lies in the obvious copy of $S_2 \times S_3$ within S_5.

(a) Show that G need not contain a product of a 2-cycle in S_2 and a 3-cycle in S_3. (Hint: Find a subgroup of $S_2 \times S_3$ with appropriate properties, then note why one can realize this as a Galois group of the indicated sort.)

(b) Show, on the other hand, that if k is a finite field, then G must indeed contain the product of a 2-cycle and a 3-cycle. (This justifies an argument Lang gives on p.274, below the middle display.)

(c) Abstracting some ideas used above, deduce the following result: If a, b are elements of a finite field k, then the polynomial $(X^3 + aX + b)(X^2 + (4a^2 + 27b^3))$ has a root in k.

***VI.2:5.** Show that the primeness assumption is needed in VI:L9 (p.322) by showing that for every composite positive integer n, the analogous statement about extensions of degree n is false.

VI.2:6. Suppose K is separable quartic extension of k (an extension of degree 4).

(a) Show that if K is *not* a quadratic extension of a quadratic extension of k, then the normal closure L of K over k has an element of degree 3.

(Suggestion: Translate the field-theoretic assumptions into conditions on $G = G(L/k)$ and on the subgroup thereof corresponding to K. Then translate these in turn into conditions on G-sets, using the result that subgroups $H \leq G$ induce transitive G-sets G/H. This will turn the question into one about subgroups of a finite symmetric group, which is not hard to prove.)

(b) Show that if $\mathrm{char}(k) \neq 2$, then the condition of (a) that K is not a quadratic extension of a quadratic extension of k is equivalent to saying that K is *not* generated over k by a root of a polynomial of the form $X^4 + aX^2 + b$ as in VI.2:1-2.

(c) Show that there exist fields K and k as in part (a).

Note on VI:L10: The words ''with $n > 2$'' at the end of the second line belong in the first line, after ''degree n''. The condition ''is the symmetric group S_n'' is ambiguous: it could mean ''acts as the full group of permutations on (the subscripts of) the n zeroes of f'', or simply ''is isomorphic, as a group, to S_n''. So for completeness, show that in the situation in question, those two conditions are equivalent, before showing that they imply the results of (a) and (b).

Note on VI:L14: I believe one is expected to assume the italicized result stated in Example 7, p.274.

Notes on VI:L16: Given a Galois extension K/k, a ''normal basis'' for K/k means a set of the form $\{\sigma\alpha\}_{\sigma \in G}$ for some $\alpha \in K$, which forms a vector-space basis of K over k. So the hypothesis that $\{\sigma\ \alpha\}_{\sigma \in G}$ is a normal basis is just equivalent to saying it is a basis. In §VI.13, Lang proves that every Galois extension has a normal basis. Though we shall not read that section, this exercise, about a property that holds assuming such a basis given, is not hard to do. Note that since α is given, when Lang says ''Show that there exists a basis ... consisting of elements of the form $S(\alpha)$'' he means ''Show that there exist subsets $S_1, \ldots, S_r$ of G such that $\{S_1(\alpha), \ldots, S_r(\alpha)\}$ forms a basis ...''.

VI.3:1. = the handout on *Quadratic Reciprocity*. Part (1)(i) is immediate from a result in Lang; you may simply cite that result. (As a week's homework, I typically assign the main parts, (1)-(8), an application such as (9), and perhaps one of (11)-(13). Part (11) requires §VI.6. (10) is important, but, together with (1)-(9), would add up to too much for a week's assignment.)

VI.5:1. Below, you will work through an easy proof of XIII:L32, a beautiful result that has been rediscovered many times. The method sketched here is adapted from H. W. Lenstra, *Inventiones math.* **25**, p.299-325, Proposition 1.3.

Let K, k and V be as in XIII:L32.

(a) Define a function $\mathrm{Tr}_G\colon V \to V$ analogous to the trace function of §VI.5, and show that this function is k-linear, and takes values in the fixed k-subspace of the action of G on V.

(b) Show that if $\varphi\colon V \to K$ is a K-linear map, and $w \in V$ has the property that $\mathrm{Tr}(Kw) \subseteq \mathrm{Ker}(\varphi)$, then $w \in \mathrm{Ker}(\varphi)$. (Hint: Write out the relations involved, and use linear independence of characters.)

(c) Deduce that $\mathrm{Tr}(V)$ lies in no proper K-subspace of V, and conclude that it contains a K-basis of V, and that this establishes the result of XIII:L32.

VI.6:1. The relation between the additive and multiplicative forms of Hilbert's Theorem 90 suggests that addition of 1 in a field of characteristic p is analogous to multiplication by a primitive pth root of unity in a field of characteristic $\neq p$. Now recall that the *discriminant* is a polynomial in n indeterminates which gets multiplied by -1 (a primitive square root of unity) on applying any odd permutation to the indeterminates. This leads to

(a) Let n be a positive integer. Deduce from results proved in Lang that there exists a rational function r_n in n indeterminates over $\mathbf{Z}/2\mathbf{Z}$ with the property that the action of any odd permutation of the indeterminates sends r_n to r_n+1. (Recall that ''rational function'' means ''element of the field $(\mathbf{Z}/2\mathbf{Z})(t_1, \dots, t_n)$'', though the elements of such a field cannot, strictly speaking, be regarded as functions.) Show, moreover, that $r_n(r_n+1)$ is a rational function of the elementary symmetric polynomials in $t_1, \dots, t_n$.

(b) Find an explicit description of such a rational function r_n for general n.

(c) For $n = 2, 3$ obtain formulas expressing $r_n(r_n+1)$ in terms of the symmetric polynomials in the given indeterminates.

VI.6:2. Let k be a field, and E the rational function field over k in $2n$ indeterminates, $x_1, \dots, x_n$, $y_1, \dots, y_n$. Let S_n act on E by simultaneously permuting the two sets of indeterminates, so that $\sigma(x_i) = x_{\sigma(i)}$, $\sigma(y_i) = y_{\sigma(i)}$, and let K be the fixed field of this action.

(a) Show that for all i, $K(x_i) = K(y_i)$. Deduce that $K(x_1, \dots, x_n) = K(y_1, \dots, y_n) = E$.

*****(b) Can you find explicit expressions for the y_i in terms of the x_i and elements of K? (I don't know whether there is an easy solution.)

Note on VI:L6: In the second sentence of part (b), for ''Let z be an element'' read ''Assume there exists an element z''. I.e., the existence of such an element is not a consequence of the preceding conditions. (If it were, the exercise would imply that every quadratic extension field was contained in a quartic extension field, which we know is not so from the case of $\mathbf{C}/\mathbf{R}$.) On the other hand, in the third line, ''Let $\alpha^2 = \gamma$'' means ''Extend E by adjoining an element α satisfying $\alpha^2 = \gamma$''.

Note on VI:L29: I suggest doing part (c) before part (b).

Note on VI:L44: Where Lang has $\lim_{n\to\infty}$, the form $\varprojlim_{n\to\infty}$ might be clearer.

Note on the next two exercises below: Neither of these actually assumes material in Lang's §VI.14 or in my comments on that section; they could be given after §VI.1. I put them here because they seemed most closely related to the ideas of §VI.14.

VI.14:1. Do VI:L26 (p.325), but give the result in as much generality as your proof naturally leads to. Also see whether you can obtain from your proof any stronger condition on the Galois groups of finite subextensions of E than that they be cyclic.

***VI.14:2.** (generalization of VI:L25, p.325). Suppose K is an infinite Galois extension of a field k, and n a positive integer such that for every finite Galois subextension $E \subseteq K$, the group $\mathrm{Gal}(E/k)$ can be generated by $\leq n$ elements. Show that k is the fixed field of some n-generator subgroup of $\mathrm{Gal}(K/k)$.

VIII.1:1. Let k be a field, K an extension field, and E, F intermediate fields. Recall that EF denotes the compositum of E and F in K.

(a) Show that $\mathrm{tr.deg.}(EF/k) + \mathrm{tr.deg.}(E \cap F/k) \leq \mathrm{tr.deg.}(E/k) + \mathrm{tr.deg.}(F/k)$.

*(b) Can the inequality of (a) be strict?

Note on VIII:L1 (p.374, based on §VIII.1): The last sentence (at least as it appears in all printings through the 4th Springer printing), ''Describe all automorphisms and their cardinality,'' is unreasonable, unless one allows a very vague sense of ''Describe''. ''Determine the cardinality of the group of all automorphisms'' would be reasonable.

Note on VIII:L6: I think that $k(x)$ and $k(y')$ in the last line should be $k_u(x)$ and $k_u(y')$.

X.4:1. Let A be a ring and $\mathbf{a} \subseteq A$ a 2-sided ideal.

(a) Show that if E is a finitely generated projective left A-module, then $E/\mathbf{a}E$ is a finitely generated projective left $A/\mathbf{a}$-module.

In the remaining parts, let us suppose $\mathbf{a}$ is contained in all maximal left ideals, and in (b)-(d), let E, F be finitely generated projective left A-modules.

(b) Show that a module homomorphism $f\colon E \to F$ is an isomorphism if and only if the induced homomorphism of $A/\mathbf{a}$-modules, $f'\colon E/\mathbf{a}E \to F/\mathbf{a}F$ is an isomorphism.

(c) Deduce that $E \cong F$ if and only if $E/\mathbf{a}E \cong F/\mathbf{a}F$.

(d) Show that E is isomorphic to a direct summand in F if and only if $E/\mathbf{a}E$ is isomorphic to a direct summand in $F/\mathbf{a}F$.

*(e) Show by example that $A/\mathbf{a}$ may have a finitely generated projective module P such that there is no finitely generated projective A-module E with $E/\mathbf{a}E \cong P$.

XIII.1:1. We observed in the comments on §XIII.1 that ${}^mR^n$ is a left module over ${}^mR^m$ and a right module over ${}^nR^n$, and that each of these rings acts on ${}^mR^n$ by endomorphisms as a module over the other.

Show that the endomorphisms of ${}^mR^n$ as a right ${}^nR^n$-module (when written to the left of their arguments and composed accordingly) form precisely the ring ${}^mR^m$, and likewise that the endomorphisms of ${}^mR^n$ as a left ${}^mR^m$-module (written to the *right* of their arguments and composed accordingly) form precisely ${}^nR^n$.

Suppose that ''by force of habit'' we write endomorphisms of the left ${}^mR^m$-module ${}^mR^n$ to the left of their arguments, and compose them accordingly. Describe the structure of the endomorphism ring in this case.

***XIII.2:1.** Let R be a field and n a positive integer. Show that any automorphism α of the ring ${}^nR^n$ preserves ranks of matrices; i.e., that if $A \in {}^nR^n$, then $\mathrm{rank}(\alpha(A)) = \mathrm{rank}(A)$. (Hint: Find some characterization of the rank of A purely in terms of the ring structure of ${}^nR^n$; e.g., in terms of the left or right ideal generated by A.)

***XIII.2:2.** Let the *inner rank* of a matrix $A \in {}^mR^n$ over a ring R be defined as in the notes to §XIII.2 in this Companion. Show by example that the inner rank of a matrix A over a commutative integral domain can be different from the rank of A over the field of fractions of R.

XIII.3:1. [cf. III.4:3 and XIII.1:1]. Let R be any ring, n any positive integer, and L any left ${}^nR^n$-module. Let $M = \{e_{11}x \mid x \in L\}$. Essentially as in Exercise III.4:3, this set may be regarded as a left

R-module. (I won't ask you to verify this this time.)

(a) Show that ${}^nM \cong L$ as left ${}^nR^n$-module. (Here we understand ${}^nM$ to be made an ${}^nR^n$-module, using the R-module structure on M, as in this section of Lang.)

This shows that every left ${}^nR^n$-module is, up to isomorphism, of the form ${}^nM$ for some left R-module M. Hence to ''understand'' ${}^nR^n$-modules, it remains only to describe their homomorphisms.

(b) Given left R-modules M and N, show that $\mathrm{Hom}_{{}^nR^n}({}^nM, {}^nN) \cong \mathrm{Hom}_R(M, N)$ as abelian groups.

(It follows that the category of left ${}^nR^n$-modules is equivalent to the category of left R-modules. Note, however, that under this equivalence, the free left ${}^nR^n$-module of rank r corresponds to the free left R-modules of rank nr. In general two rings having equivalent module-categories are called *Morita equivalent*. There is a general theory describing how this happens; the above is the simplest nontrivial case.)

***XIII.4:1.** This problem concerns a possible converse to XIII:L26 (p.548).

Suppose n is a positive integer, and

$$f \in \mathbf{Z}[t_{11}, t_{12}, \ldots, t_{nn}]$$

a polynomial in n^2 indeterminates, and suppose f has the property that for every commutative ring A, every n-generator ideal $I = Ax_1 + \ldots + Ax_n \subseteq A$, and every family of n^2 elements $c_{ij} \in A$, if we define $y_i = \Sigma\, c_{ij} x_j$ and $I' = Ay_1 + \ldots + Ay_n$, then we have

$$f(c_{11}, c_{12}, \ldots, c_{nn})\, I \subseteq I'.$$

Must f be divisible (as a polynomial) by $\det(t_{ij})$?

XIII.4:2. Let m and n be positive integers, and let E be a free module of rank m over a commutative ring R. Show that $L_a^n(E)$ is a free R-module, and determine its rank. (For $n = m$ you should, of course, get Corollary XIII.4.12, p.517.)

XIII.4:3. (i) = XIII:L25 (p.547). Hint: The $n = 2$ case is easy. To simplify the problem for general n, apply appropriate invertible transformations to pairs of columns.

*****(ii) More generally, suppose we are given an $m \times n$ matrix A over a principal ideal domain, with $m < n$. Can you find necessary and sufficient conditions for this to be completable to an $n \times n$ matrix of determinant 1?

XIII.5:1. We have been studying the classes of symmetric and alternating bilinear forms on a free module. This exercise will examine the question of whether there are any other ''natural'' classes of bilinear forms. For simplicity, we will assume our base ring is a field.

Given a field k, let us define a ''subfunctor X of L^2'' to mean a construction associating to every finite-dimensional k-vector-space V a subspace $X(V) \subseteq L^2(V, k)$ in such a way that for every linear map $f: V \to W$ among vector spaces, the induced map $L^2(f): L^2(W) \to L^2(V)$ carries $X(W)$ into $X(V)$. Let us partially order the subfunctors of L^2 by writing $X \leq X'$ if for all V, one has $X(V) \subseteq X'(V)$.

Show that the subfunctors of L^2 are precisely L^2 itself, L_a^2, L_s^2, and 0, and determine the partial ordering on this set of four subfunctors. This ordering will depend (to a slight extent) on k.

(Remark: What we are calling $L^2(V)$ can be naturally identified with $\mathrm{Hom}_k(V \otimes_k V, k)$. A more natural form of the above question would ask for all subfunctors of the functor $V \mapsto V \otimes_k V$. The above results would follow from the answer to this question with the help of the standard correspondence between subspaces of a finite-dimensional vector space and subspaces of its dual. But I forewent this elegance to fit the question to the reading. Note also that one might generalize the original question by replacing ''bilinear'' with ''n-linear'' for fixed $n \geq 0$; equivalently, in the more natural form just mentioned, one might replace the 2-fold tensor product with the n-fold tensor product. The answer then becomes considerably more complicated!)

XIII.5:2. Let k be a field, and κ an infinite cardinal.

(a) Prove that the dual of a κ-dimensional k-vector-space has dimension at least $\max(2^\kappa, \text{card } k)$.

Suggestion: If $\text{card } k < 2^\kappa$, look at the cardinality of the dual space. In the contrary case, show that the dual has a subspace isomorphic to the space k^ω of all sequences of elements of k, and show that the exponential sequences $(1, c, c^2, \dots)$ $(c \in k)$ are linearly independent.

The dual space considered in the above result is clearly isomorphic to the k-vector-space of all k-valued functions on a set of cardinality κ. It is interesting that one can get a similar result for a much more restricted space of functions:

(b) Let S be a set of cardinality κ, and V the space of k-valued functions on S that assume only finitely many distinct values. (Equivalently, the space spanned by the characteristic functions of all subsets of S.) Show that V has dimension 2^κ.

Suggestion for showing ''$\geq$'': Let X be a set of cardinality κ, and note why you can assume without loss of generality that S is the set of all finite subsets of X. For every subset $Y \subseteq X$, let $\varphi_Y \in V$ be the function which takes the value 1 at members of S which are subsets of Y, and 0 at all others. Show that these 2^κ members of V are k-linearly independent.

***XIV.2:1.** (Cf. Corollary XIV.2.3, p.558.) Let $A_1, \dots, A_r, B_1, \dots, B_r$ be $n \times n$ matrices over a field k. Assume there is an invertible matrix C' over an extension field k' of k such that $B_i = C' A_i C'^{-1}$ $(i = 1, \dots, r)$. Must there exist an invertible matrix C over k itself with this property? If you can't answer the question in complete generality, you might see whether you can do so under some additional assumption.

XIV.3:1. (a) = XIV:L10 (p.568). (Note that while a polynomial $f(t) \in k[t]$ can be evaluated at any element of k, we can merely say that a rational function $r(t) = p(t)/q(t) \in k(t)$ can be evaluated at ''almost all'' elements of k, namely, those that are not roots of q, assuming our fraction written in lowest terms. Likewise, one can only say that $r(A)$ can be evaluated at ''most'' matrices. You should describe the sense in which one can speak of evaluating $r(A)$, and then state and prove the result asked for.)

(b) Suppose M is the matrix of an endomorphism f of a finite-dimensional vector-space V, which again satisfies the hypothesis of Theorem XIV.3.10. We know that f induces endomorphisms $L^2(f)$, $L^2_s(f)$ and $L^2_a(f)$ of the spaces $L^2(V)$, $L^2_s(V)$ and $L^2_a(V)$ respectively. Determine the characteristic polynomials of these maps.

XVI.1:1. (a) Show that if A is a nonzero finitely generated abelian group, then the tensor product $A \otimes_{\mathbf{Z}} A$ is nonzero.

In contrast:

(b) Let $p \in \mathbf{Z}$ be a prime, and let the group Z_{p^∞} be defined as in I.8:3. Show that $Z_{p^\infty} \otimes_{\mathbf{Z}} Z_{p^\infty} = \{0\}$.

*****(c) Prove the analog of (a) for finitely generated modules over an arbitrary commutative ring. (Whether this part should be ''starred'' really depends on what previous chapters have been covered. But the next one certainly deserves its star.)

*****(d) Does there exist a commutative ring k and a k-module M such that $M \otimes_k M$ is nonzero, but $M \otimes_k M \otimes_k M = \{0\}$? If you can't answer this for general k, can you answer it for $k = \mathbf{Z}$?

XVI.1:3. Let E and F be vector-spaces over a field k, and z any element of $E \otimes_k F$.

(a) Show that one can write $z = \Sigma_{i=1}^{r} x_i \otimes y_i$ where $x_1, \dots, x_r$ are linearly independent elements of E, $y_1, \dots, y_r$ are linearly independent elements of F, and $r \geq 0$.

(b) Suppose we have another expression for the same element, $z = \Sigma_{i=1}^{r'} x'_i \otimes y'_i$, where the x'''s and y'''s again form linearly independent families. Show that $r' = r$, that the space spanned by the x'''s is the same as the space spanned by the x's, and that the space spanned by the y'''s is the same as the space

spanned by the y's. In fact, show that if we write $\mathbf{x}$ for the row vector of elements of E formed by the x's, $\mathbf{x}'$ for the row vector formed from the x''s, and likewise $\mathbf{y}$ and $\mathbf{y}'$ for the column vectors of elements of F formed from the y's and the y''s, then there exists an invertible $r\times r$ matrix U over k such that $\mathbf{x}' = \mathbf{x}U$, $\mathbf{y}' = U^{-1}\mathbf{y}$.

XVI.1:4. Let n be a positive integer and k a commutative ring, and let $\otimes^n$ denote the functor from k-modules to k-modules that takes every k-module M to the n-fold tensor product $M \otimes_k \dots \otimes_k M$, which we shall write $\otimes^n M$. Given any homomorphism $h\colon M \to N$ of k-modules, the induced homomorphism of n-fold tensor products will be written $\otimes^n h\colon \otimes^n M \to \otimes^n N$.

By an *endomorphism f of the functor* $\otimes^n$ we shall mean any construction associating to each k-module M an endomorphism f_M of the k-module $\otimes^n M$, which respects homomorphisms of k-modules, in the sense that for every such homomorphism $h\colon M \to N$, the following diagram commutes:

$$\begin{array}{ccc} \otimes^n M & \xrightarrow{\otimes^n h} & \otimes^n N \\ \Big\downarrow f(M) & & \Big\downarrow f(N) \\ \otimes^n M & \xrightarrow{\otimes^n h} & \otimes^n N \end{array}$$

One defines the sum of two endomorphisms of $\otimes^n$, and the result of multiplying such an endomorphism by an element of k, in the obvious ways, and one thus sees that such endomorphisms form a k-module (at least, if we ignore the question of whether they form a set!).

As an example, it is easy to see that for each element σ of the symmetric group S_n, there is an endomorphism $\bar\sigma$ of $\otimes^n$ whose action on decomposable elements is given by $\bar\sigma(x_1 \otimes \dots \otimes x_n) = x_{\sigma^{-1}(1)} \otimes \dots \otimes x_{\sigma^{-1}(n)}$. We can now state the problem:

Show that every endomorphism f of $\otimes^n$ is a k-linear combination of the endomorphisms $\bar\sigma$ ($\sigma \in S_n$). (Suggestion: Look at the behavior of f on an appropriate ''universal'' element.)

Remark: Clearly one can also compose these endomorphisms, hence they have a natural structure of *k-algebra*. It is not hard to deduce from the above result that this algebra is isomorphic to the group algebra kS_n. But I will spare you the writing out, and myself the grading, of the verification of this fact.

XVI.1:5. [Assumes the material on *Tensor products in the noncommutative context*, this Companion.]

Let k be a field, and V a 1-dimensional k-vector-space, with basis $\{x\}$.

(i) Determine all ways to make V into a $(k[t], k[t])$-bimodule so that the right and left actions of elements of k agree with their actions under the given k-vector-space structure.

(ii) For every pair of bimodules E and F of the sort characterized in (i), determine the $(k[t], k[t])$-bimodule structure of $E \otimes_{k[t]} F$.

(iii) Show that for appropriate choice of E and F as above, one can on the one hand get $E \otimes_{k[t]} F$ nonisomorphic as a k-vector-space to $F \otimes_{k[t]} E$, and on the other hand, get these objects isomorphic as k-vector-spaces but nonisomorphic as bimodules.

XVI.2:1. Here are a few examples showing that the homomorphism of Proposition XVI.2.5 (p.610) can fail to be an isomorphism if E and F are not taken free of finite rank.

(i) Show that if R is a field, and E, F are infinite-dimensional vector spaces over R, then that homomorphism is one-to-one, but not onto.

(ii) Show that if $R = \mathbf{Z}$, $E = Z_{p^\infty}$ (see I.8:3), and $F = \mathbf{Q}$, then that homomorphism is onto but not one-to-one. (Hint: Show that $\mathrm{End}(Z_{p^\infty})$ is torsion-free as an additive group, and use this to deduce that the tensor-product of endomorphism rings is nontrivial.)

(iii) Show that if $R = \mathbf{Z}$, $E = Z_{p^2} \oplus Z_p$, and $F = Z_p$, then the homomorphism is neither one-to-one nor onto.

On the other hand, there are some nice cases not covered by the proposition in which one can show that the homomorphism is an isomorphism:

(iv) Show that if E *or* F is free of finite rank, then that homomorphism is an isomorphism. (Hint: In that case, you get a decomposition of $E \otimes F$ as a direct sum of copies of F, and can use the ideas in the comment regarding p.505, bottom, in this Companion about using matrices to represent direct homomorphisms among direct sums of modules.)

(v) Show, still more generally, that the homomorphism is an isomorphism whenever E or F is a finitely generated projective module.

XVI.3:1. (i) Suppose R is a ring and G a functor from left R-modules to abelian groups, which respects the group structure on homomorphisms; i.e., such that given $f, g: E \to E'$, we have $G(f+g) = G(f) + G(g)$. Show the equivalence of conditions **F1** and **F2** (pp.612-613) with the functor G in place of the functor $F \otimes -$ of those pages.

(ii) Show by example that these two conditions are not equivalent to **F3** in this generality, by looking at the functor $L(F, -)$ for an appropriate left R-module F. (Suggestion: Figure out to what extent such a functor *must* preserve exactness, then look for an example where the case that need *not* hold fails.)

Remark: Along with functors of the forms $F \otimes -$ and $L(F, -)$, there is one other familiar class of functors on categories of modules that respects the group structure on morphisms: the contravariant functors $L(-, E)$. Functors of each of these sorts in general preserve exactness at most, but not all parts of a short exact sequence. The F or E for which they do preserve exact sequences (in which case they are called *exact functors*) are called, respectively, *flat* modules (discussed in this section), *projective* modules (pp.137-139) and *injective modules* (§XX.4).

XVI.3:2. Prove the following partial converse to Proposition XVI.3.2: If R is a unique factorization domain, and every torsion-free R-module is flat, then R is a principal ideal domain.

(Suggestion: Reduce the problem to showing that if $x, y \in R$ have g.c.d. 1, then $xR + yR = R$, and then somehow get this equality from the fact that $xR + yR$ is flat. One way to use flatness might be to apply Lemma XVI.3.3 with $M \to F$ the natural map $xR \oplus yR \to xR + yR$, and $E = R/(xR + yR)$.)

***XVI.3:3.** This exercise will prove the implication (c)⇒(a) of Theorem XVI.3.2 by an argument close to that of Lazard's original paper.

Let E be a right R-module satisfying condition (b) of that theorem. Let H be a free right R-module on a basis $\mathfrak{B}$, given with a homomorphism $h: H \to E$ such that for every $x \in E$, infinitely many elements of $\mathfrak{B}$ are mapped to x. (For instance, we could let F be the free module on the product of the set of natural numbers and the underlying set of E, and map this to E in the obvious way.)

Now let I be the set of ordered pairs (H_i, K_i) where H_i is the free submodule of H generated by a finite subset of $\mathfrak{B}$, and K_i is a submodule of $H_i \cap \mathrm{Ker}(h)$ such that H_i/K_i is also free. For $i = (H_i, K_i) \in I$ we shall write $F_i = H_i/K_i$. Note that for each $i \in I$, the map h induces a homomorphism $h_i: F_i \to E$. Given $i = (H_i, K_i)$, $j = (H_j, K_j) \in I$, we shall write $i \leq j$ if $H_i \subseteq H_j$ and $K_i \subseteq K_j$.

(i) If $i \leq j \in I$, describe a natural homomorphism $f_{j,i}: F_i \to F_j$ over E, and show that if $i \leq j \leq k \in I$, then $f_{k,j}\, f_{j,i} = f_{k,i}$.

(ii) Show that if $i \in I$, and F is a free module of finite rank given with a homomorphism into E, and $f: F_i \to F$ is any homomorphism over E, then there exists $j \geq i$ such that F_j is isomorphic to F as a module with a homomorphism from F_i into it, and a homomorphism from it into E. (I.e., there is an isomorphism $F_j \cong F$ which makes commuting triangles with the maps out of F_i and into E.) This proof will make use of the assumption that infinitely many elements of $\mathfrak{B}$ map to each element of E, but the assumption that E satisfies (b) will not yet be needed.

(iii) Show by combining the hypothesis (b) on E and the result of (ii) that the partially ordered set I is

directed, and that $\varinjlim_i F_i \cong E$.

XVI.3:4. Let R be a commutative ring and $\mathbf{m}$ a maximal ideal of R. Thus, $R/\mathbf{m}$ is a field, and $\mathbf{m}/\mathbf{m}^2$ can be regarded as a vector-space over that field.

(i) Show that if this vector-space has dimension > 1, then $\mathbf{m}$ is not flat as an R-module.

(ii) Show that all the examples of non-flat ideals given in the comment to p.614 (top two paragraphs) in this Companion are instances of this result.

XVI.3:5. (a) Suppose $0 \to N \to M \to F \to 0$ is a short exact sequence of right modules over a ring R, and $f\colon F_0 \to F$ a module homomorphism. Show how to obtain from this map a module M_0, a short exact sequence $0 \to N \to M_0 \to F_0 \to 0$, and a map of short exact sequences

$$\begin{array}{ccccccc} 0 & \to & N & \to & M_0 & \to & F_0 & \to & 0 \\ & & \downarrow{=} & & \downarrow & & \downarrow f_0 & & \\ 0 & \to & N & \to & M & \to & F & \to & 0. \end{array}$$

(b) Verify that your construction is functorial; that is, that given maps from two free modules F_0 and F_1 into F, and a homomorphism $F_0 \to F_1$ respecting the maps into F, you get a homomorphism between your short exact sequences that respect your maps into the given sequences.

(c) Verify that your construction respects direct limits.

(d) Combining these observations with Lazard's Theorem, obtain a proof of Lemma XVI.3.3 (p.614) based on the fact that the desired result holds when F is projective (and so in particular, when it is free).

(e) Similarly, prove the first assertion of Proposition XVI.3.4 (p.615) by first establishing it when F'' is free, then applying (a)-(c).

XVI.3:6. Let $0 \to E' \to E \to E'' \to 0$ be a short exact sequence of left modules over a ring R, and F a right R-module. If F is E'-flat and E''-flat, must it be E-flat?

XVI.3:7. (Cf. XVI:L6, p.638.) Suppose E is a flat right R-module and F is an (R,S)-bimodule which is flat as a right S-module. Show that $E \otimes_R F$ is flat as a right S-module.

XVI.3:8. Let R be the ring of all continuous functions on the interval [0,1], and E the R-module of all ''germs at 0'' of such functions; i.e., the R-module R/I, where I is the ideal of all functions zero in the neighborhood of 0. Show that E is a flat R-module. (This same ring and ideal appear in the more difficult Exercise III.4:2.)

XVI.3:9. Suppose S is a commutative integral domain, K is its field of fractions, and R is a subring of S having the same field of fractions.

(a) Show that every R-module homomorphism of S into R is given by multiplication by a member of R, and that the set of $r \in R$ such that multiplication by r carries S into R forms an ideal of R, which is the largest ideal of R that is also an ideal of S. This is called the *conductor ideal* of R in S.

(b) Determine the conductor ideal of $k[x]$ in $k[x, x^{-1}]$, and the conductor ideal of $k[x^2, x^3]$ in $k[x]$. Under what conditions on R and S will 1 belong to the conductor ideal?

(c) Show that if the identity map of S, regarded as a homomorphism of R-modules, factors through a free R-module, then $S = R$.

(d) Deduce that if S is finitely presented as an R-module (in particular, if R is Noetherian and S is finitely generated as an R-module) and is strictly larger than R, then S is not flat as an R-module.

XVI.3:10. Let k be a field, and $R = k\langle x, y\rangle$ the free associative k-algebra on indeterminates x and y, equivalently, the monoid algebra over k on the free monoid on x and y.

(a) Let P denote the right R-module presented by two generators α and β and the one relation

$\alpha x - \beta y = 0$. How that $\mathrm{Hom}_R(P, R) = 0$.

(b) Deduce that if $f: R \to S$ is a ring homomorphism such that either $f(x)$ or $f(y)$ is invertible in S, then S is not flat as a right R-module (under the induced module structure $s \cdot r = s\, f(r)$).

In particular, this shows that R cannot be embedded in a division ring that is flat as a right R-module, and also that the group algebra of the free group on x and y is not flat as a right R-module.

XVI.3:11. Within the division ring of quaternions, let

$R = \{a + bi + cj + dk \mid a, b, c, d \in \mathbf{Z}\}$,

$S = \{(a + bi + cj + dk)/2 \mid a, b, c, d \in \mathbf{Z},\ a \equiv b \equiv c \equiv d \pmod 2\}$,

$K = \{a + bi + cj + dk \mid a, b, c, d \in \mathbf{Q}\}$.

Clearly, R is a ring and K a division ring.

(a) Verify that S is also a ring.

(b) Show that K is flat both as a right R-module and as a right S-module.

(c) Show that S is not flat as a right R-module. (Suggestion: Look for analogs of the ideas of XVI.3:9.)

(d) Show that every right ideal of S is principal, and deduce that every right S-module which is torsion-free as an abelian group is flat.

Remark: From (c) and (d), we see that S is ''better behaved'' than R; e.g., S is a ''noncommutative principal right (and left) ideal domain'', while R is not. So S is commonly called the ''ring of integer quaternions'', while there is no special name for R.

***XVI.3:12.** (a) Show that if R is a commutative Noetherian ring, and P a finitely presented R-module, then there exists a homomorphism h of P into a free R-module F of finite rank such that any homomorphism of P into a free R-module factors through h.

(b) Deduce that the direct product of an (in general, infinite) family of flat R-modules over a commutative Noetherian ring is flat. (We already know this for direct products of finite families, since these are the same as direct sums.)

(c) To show that (b) fails for non-Noetherian R, let G be a monoid isomorphic to the monoid $[0, \infty)$ of nonnegative real numbers, but written multiplicatively as $\{x^r \mid r \in [0, \infty)\}$, let S be the monoid algebra kG, and let R be the factor-ring of S by the ideal generated by all x^r with $r > 1$. (One might write this sloppily as $k[x^{[0,1]}]$.) Show that a direct product of countably many copies of R is not flat. (Hint: The element $a = (x, x^{1/2}, x^{1/4}, \dots)$ satisfies the relation $ax = 0$.)

(d) Prove versions of (a) and (b) for noncommutative R.

XVI.4:1. Let $f: R \to S$ be a ring homomorphism, let A be a right R-module, and let B be a left S-module. Establish a natural isomorphism $A \otimes_R B = A_S \otimes_S B$ as abelian groups. Describe the isomorphism and its inverse explicitly on generators of these tensor products if your method of obtaining the isomorphism does not do so.

XVI.5:1. Let R be a commutative ring, and E and F R-modules. It is easy to see that the construction of Lang's Corollary XVI.5.5 (p.628) gives a functorial homomorphism $\varphi_{E,F}: E^\vee \otimes F \to \mathrm{Hom}(E, F)$, without the assumption of that corollary that the modules be free of finite rank.

Characterize some of the following classes of modules. (Instructor should specify how many; e.g., ''two''.)

(a) Those E such that for all F, $\varphi_{E,F}$ is an isomorphism.

(b) Those F such that for all E, $\varphi_{E,F}$ is an isomorphism.

(c) Those F such that for all free modules E, $\varphi_{E,F}$ is an isomorphism.

(d), (e), ... Similar cases with ''an isomorphism'' replaced by ''one-to-one'' or ''onto'', and/or with other classes of modules, so long as you don't hand in two results with slightly varying hypotheses but

essentially the same proofs.

XVI.6:1. Let A be a $\mathbf{Z}$-graded or $\mathbf{Z}/2\mathbf{Z}$-graded algebra, as in the last paragraph on p.631 of Lang.

(i) If the grading monoid is $\mathbf{Z}$, show that what Lang calls the ''super algebra A_{su}'' is actually isomorphic to the ''ordinary algebra A''.

(Suggestion: Take a polynomial ring $R[x]$, grade it so that x has degree 1, and determine the unique R-algebra isomorphism $R[x] \to R[x]_{su}$ carrying x to x. This will involve multiplying homogeneous elements of each degree by an appropriate coefficient. Verify that these same coefficients give an isomorphism in the case of any $\mathbf{Z}$-graded algebra.)

(ii) On the other hand, show that if we grade the complex numbers by $\mathbf{Z}/2\mathbf{Z}$, letting $\mathbf{C}_0 = \mathbf{R}$ and $\mathbf{C}_1 = i\mathbf{R}$, then $\mathbf{C}_{su}$ is not isomorphic to $\mathbf{C}$.

***XVI.8:1.** For E a module over a commutative ring R and n a positive integer, does $S^n E = 0 \Rightarrow \otimes^n E = 0$?

(This question was asked by Moss Sweedler. I succeeded in answering it a few years ago; but it has potential generalizations in the context of the representation theory of the symmetric group which could form an interesting research topic.)

XX.4:1. *(a) = XX:L20 (p.830). (Suggestion [> XX:L23]: Show that any ideal of $S^{-1}A$ has the form $S^{-1}J$ for an ideal J of A. Use the fact that J can be finitely presented as an A-module to show that any $S^{-1}A$-module homomorphism $f: S^{-1}J \to S^{-1}I$ is of the form $s^{-1}\bar{\varphi}$, where $s \in S$, and $\bar{\varphi}: S^{-1}J \to S^{-1}I$ is the map induced by an A-module homomorphism $\varphi: J \to I$. Then get the desired result using XX:L23.)

Parts (b) and (c) below refer to the statement of the exercise comprising part (a), but do not depend on it or one another. Part (d) depends on (a) and (c).

(b) (T.Y.Lam) Show that the conclusion of the above exercise remains true if the assumption ''A is Noetherian'' is replaced by the condition that the module I has no S-torsion; i.e., that no nonzero element of I is annihilated by an element of S.

***(c)** Show that the result in question becomes false if neither the Noetherian condition of part (a) nor the no S-torsion condition of (b) is assumed. (Suggestion [> XX:L21(a)]: Let k be an algebraically closed field, and regard the polynomial ring $k[x,y]$ as a subring of the ring of k-valued functions in two k-valued variables, $k^{k\times k}$. Let A be the subring of $k^{k\times k}$ spanned by $k[x,y]$ and all functions of finite support, and let $I = k^{k\times k}$ regarded as an A-module. XX:L21(a) reduces the problem of showing I injective to the same problem for certain much smaller modules; get these cases with the help of the fact that the element of A having the value 1 at a single point and 0 elsewhere is idempotent, hence induces a direct product decomposition of A as a ring. Now let $S = k[x] - \{0\} \subseteq k[x,y] \subseteq A$, and get descriptions of $S^{-1}A$, and $S^{-1}I$. Show that if $S^{-1}I$ were injective, then multiplication by y would map it *onto* itself, but that it does not.)

***(d)** [> §III.10] Give a version of the proof of (a) above that proceeds via the following three results: (i) if A is Noetherian, then a direct limit of injective A-modules is injective; (ii) for any commutative ring A, any A-module I, and any multiplicative set S, $S^{-1}I$ can be expressed as a direct limit of copies of I; (iii) if an $S^{-1}A$-module is injective as an A-module, it is injective as an $S^{-1}A$-module.

Deduce from (c) that the condition that A be Noetherian cannot be dropped from (i) above.

Notes on exercises connected with Lang's Appendix A2: Exercises I.1:1 above assumes that Appendix A2.1 was read before §I.1. If they are read in the reverse order, that can be considered an exercise for Appendix A2.1.

Exercises A2.2:1-4 below concern Zorn's Lemma, developed in Appendix 2.2 and in my handout on ''The Axiom of Choice, Zorn's Lemma and all that''. Since I introduce this material between sections I.6

and I.7 of Lang, most of the exercises below assume the material in the first part of Chapter I.

If Exercise A2.2:1 is not assigned, it should probably be discussed in class.

A2.2:1. Show the following facts. (Negative results should be shown by constructing counterexamples, and proving relevant properties of these if they are not obvious.)

(a) Every nontrivial finitely generated group G has a maximal proper subgroup. (This means a maximal element in the class of proper subgroups, partially ordered by inclusion.)

(b) Every nontrivial finitely generated group G has a maximal proper *normal* subgroup (a maximal member of the class of proper normal subgroups).

(c) A nontrivial finitely generated group G need not have a minimal nontrivial subgroup. (I.e., there exists a nontrivial finitely generated group G not having such a subgroup.)

(d) Every nontrivial finitely generated group G has a simple homomorphic image.

(e) A maximal proper subgroup of a group G need not be normal, and a maximal proper normal subgroup of G need not be a maximal proper subgroup of G.

(f) A nontrivial *non*-finitely-generated group G need not have a maximal proper subgroup, nor a maximal proper normal subgroup, nor a simple homomorphic image.

(Remark: The terms ''maximal proper subgroup'', ''minimal nontrivial subgroup'', etc. are often shortened to ''maximal subgroup'', ''minimal subgroup'', etc.. That is, since the properties of maximality and minimality in the class of *all* subgroups of G characterize subgroups that can be named much more easily, it is taken for granted that when one refers to maximality or minimality, one must be thinking of proper, respectively nontrivial, subgroups. But there are situations where such abbreviation can lead to confusion; so be cautious in using them. If you are going to be writing extensively about maximal proper subgroups and want to use this abbreviated term, you should begin by *saying* that ''maximal subgroup'' will be understood to mean ''maximal proper subgroup''.)

A2.2:2. [cf. A2.2:1]. (a) Does every nontrivial finitely generated semigroup have a simple homomorphic image? (A semigroup S is called *simple* if it is nontrivial, and every homomorphism from S to another semigroup is either an isomorphism, or is trivial, i.e., has one-element image.)

(b) If G is a group, will every nontrivial finitely generated G-set have a simple homomorphic image?

A2.2:3. Let X be a set. If $\leq$ and $\leq'$ are two partial orderings on X, we say that $\leq'$ is a *refinement* of $\leq$ if

$$(\forall\, x, y \in X)\ \ x \leq y \ \Rightarrow\ x \leq' y,$$

i.e., if, when we regard the relations $\leq$ and $\leq'$ as subsets of $X \times X$, the former is contained in the latter. Clearly, ''is a refinement of'' is a partial ordering on the set of partial orderings on X!

(a) Show that every partial ordering on X can be refined to a partial ordering *maximal* with respect to refinement.

(b) Show that every such maximal partial ordering is a total ordering.

(c) Is every partial ordering on X the *intersection* of some set of total orderings?

A2.2:4. For this exercise, we need the following definition and observations.

A *subdirect product* of a family of groups G_i $(i \in I)$ means a subgroup G of the direct product $\prod_I G_i$, whose projection to each G_i is *surjective*. (Precisely: the composite of the inclusion $G \to \prod_I G_i$ with each projection $\prod_I G_i \to G_j$ is surjective.) Given a group G, we say that G ''can be written as a subdirect product of'' a family of groups G_i if it is isomorphic to such a subgroup of their product. Note that in this situation, each G_i will be isomorphic to a factor-group G/N_i, and the intersection of these N_i will be $\{e\}$. Conversely, given G and any system of normal subgroups with trivial intersection, we get a representation of G as a subdirect product of the factor-groups G/N_i.

A group G is called *subdirectly irreducible* if the only representations of G as a subdirect product are those in which at least one of the maps $G \to G_i$ is an isomorphism.

(a) Show that G is subdirectly irreducible if and only if it has a *smallest nontrivial* normal subgroup (a nontrivial normal subgroup contained in all nontrivial normal subgroups).

(b) Show that if G is a group, x an element of G, and N a normal subgroup of G maximal for the property $x \notin N$, then G/N is subdirectly irreducible.

(c) Deduce that every group can be written as a subdirect product of a family of subdirectly irreducible groups.

(Exercise I.6:5 concerned some restricted cases of the above results. There, finiteness allowed one to work without Zorn's Lemma, and to express the conclusion in terms of embedding in a product of groups of smaller order.)

A2.2:5. Let G be a group. By a ''maximal abelian subgroup'' of G we understand a maximal member of the class of abelian subgroups; not necessarily maximal as a subgroup of G. Recall also that the *centralizer* of a subset $X \subseteq G$ means $\{g \in G \mid (\forall\, x \in X)\ xg = gx\}$.

(a) Show that every abelian subgroup of G is contained in a maximal abelian subgroup.

(b) Show that the maximal abelian subgroups of G are precisely those subgroups of G that are their own centralizers.

(c) Show that for a subgroup (in fact, a subset) A of G, the following three conditions are equivalent: (i) A is the intersection of a nonempty set of maximal abelian subgroups of G. (ii) A is the centralizer of some subset of G containing A. (iii) A is an abelian subgroup of G, and for every element $g \in G$ not in A, there is an element $h \in G$ which commutes with all elements of A, but not with g. (Note: the next part, (d) is an easy consequence of (c). But if you are having trouble proving one of the implications needed for (c), you might try to work out a direct proof of (d), then see whether the argument you arrive at can be modified to complete your proof of (c).)

(d) Show that the *center* of G is the intersection of all maximal abelian subgroups of G.

A2.4:1. Let A be a partially ordered set. Show that the partial ordering of A can be refined (in the sense of A2.2:3) to a well-ordering if and only if every nonempty subset of A has an element minimal with respect to the given ordering. (Hint: Mimic the proof of Theorem A2.4:1.)

Note on A2:L13 (i.e., Exercise 13 at the end of Appendix 2, p.893). Hint: The set of maps from X to Y is often denoted Y^X. Think ''laws of exponents'', then try to justify the argument precisely.

INDEX TO DISCUSSION OF LANG'S EXERCISES IN THIS *COMPANION*
mostly in the above collection of exercises, but occasionally elsewhere
(excluding errata that apply only to earlier printings, for which see ''Errata'' section)

Exercises to Chapter I, pp.75-82: #3 clarified in I.3:1. Note on #4 at end of exercises on §I.3. #5 discussed and generalized in I.3:2. #6 incorporated into I.6:4(a). #7 incorporated into I.5:1. #15 and #16 incorporated into I.5:4. #24 incorporated into I.6:1. #38(d) generalized in I.5:7. In #44 an assumption can be deleted, as noted at end of exercises on §I.2. #47 incorporated into I.5:6. #57: Examples at end clarified at end of exercises on §I.5.

Exercises to Chapter II, pp.114-116: #1 subsumed by II.2:1. #3 incorporated into II.4:2. #5 clarified at end of exercises on §II.5. #9 belongs in Chapter IV but done in Companion, as noted at end of exercises on §II.1. Clarification of #12 and cross-reference from #13-19 to III:L11-13 noted at end of exercises on §II.1

Exercises to Chapter III, pp.165-171: Hint to #10 given at end of exercises on §III.1. Note that #5-#8 can go with §I.7 after exercises for that section. #6 **corrected** after exercises on §III.5. #14 clarified after

exercises on §III.7. #16 discussed (''trivial or misstated'') after exercises on §III.9; related questions given as III.10:1-2.

Exercises to Chapter IV, pp.213-220: #8 and #9 incorporated into II.3:1. Notation of #12 clarified at end of exercises on §IV.6.

Exercises to Chapter V, pp.253-259: Comments on #17 and #21, and hint to #28, given after exercises for this chapter. #24 is covered in discussion *Examples involving inseparable extensions* (this Companion, note to p.252 of Lang).

Exercises to Chapter VI, pp.320-332: #6 clarified in discussion following exercises to §VI.6. #8 generalized and discussed in VI.2:1. Converse to #9 given as VI.2:5. #10, #14, #16 clarified following exercises for VI.2. #25 generalized in VI.14:2. Strengthened version of #26 asked for in VI.14:1. Suggestion on #29 following exercises to §VI.6.

Exercises to Chapter VIII, pp.374-375: #1 clarified/criticized after exercises on §VIII.1. #6 **corrected** after exercises on §VIII.1.

Exercises to Chapter XIII, pp.545-552: #25 incorporated into XIII.4:3. Possible converse to #26 given as XIII.4:1. #32, with different approach, given as VI.5:1; still another approach noted in *Further observations on linear independence of characters* (this Companion, note to p.284 of Lang).

Exercises to Chapter XIV, pp.567-570: #10 incorporated, with clarification, in XIV.3:1.

Exercises to Chapter XVI, pp.637-639: #6 generalized to bimodules in XVI.3:7. #13 developed in detail (with some **corrections** noted) in *Further criteria for flatness* (this Companion, note to p.613 of Lang), and alternative argument to one implication given as XVI.3:3.

Exercises to Chapter XX, pp.826-832: #20 incorporated into XX.4:1.

Exercises to Appendix 2, pp.892-893: Hint to #13 given after exercises to that Appendix.

ERRATA TO EARLIER PRINTINGS OF LANG'S **ALGEBRA** (3RD EDITION)

and minor errata to the current printing

This list does not include corrections noted in the main part of this ''Companion'', i.e., corrections to the latest printing that seemed important enough to bring to the student's attention. Thus, only students who have earlier printings need to use it. (However, students with the current printing who find a minor typo in their copy of Lang might look below before reporting it, to see whether I had noted it but felt it not sufficiently important to record in the main part of this ''Companion''.)

Below, I will denote the six Addison-Wesley printings by 1a-6a, and the printings of the revised Springer version by 1s, 2s, etc.. (The most recent printing I have, which I am using to update this list, is 4s, though I know that 5s has come out.)

In the code in brackets before each item, numbers in boldface indicate printings to which an erratum applies, while italic numbers are printings to which it does not apply. Thus [**2a**|*4a*] indicates an error found in the 2nd Addison-Wesley printing, and corrected by the 4th; in such cases, it probably also occurred in the 1st printing, and not in printings after the 4th, but I don't know whether it occurred in the 3rd printing. Minor errors still occurring in the latest printing that I have are marked [**4s**]. Occasionally, an error first appears in a later printing; in that case I may have an indication like [*6a*|**1s**], meaning that the error was absent in (and presumably up to) the 6th Addison-Wesley printing, but appeared in the first Springer printing, and has not been corrected in the most recent printing that I have.

I am indebted to students in my past Math 250 classes, especially David Gay (Fall 1993), Chee-Whye Chin (Fall 1995) David Wasserman (Spring 1997) and Steven Sam (Fall 2006) for bringing to my attention many of these errors. While Lang was alive, I would point them out to him, and he generally corrected them; I now similarly communicate them to the publishers. My colleagues Paul Vojta and Bjorn Poonen and students in their courses have also contributed a number of corrections.

Along with errors, I note a few additions of material (mostly new exercises) that have been made in recent printings. Since information about material first added in (say) printing 1s should be of interest to the same set of readers (those having earlier printings) as an error first corrected there, I will mark such cases in the same way, as [**6a**|*1s*]. In such cases I will usually not give full details of the new material; the student interested in those details should find a copy of the relevant printing.

[*6a*|**1s**|*4s*] P.viii, top line of last paragraph: ''Lindeman'' should have a double n.

[**3a**|*4a*] P.xii. The page number for Chapter IV, §4 should be ''186'', and the page number for Chapter VI, §3 should be ''276''.

[**6a**|*1s*] P.xii: The beginning of the title of Chapter IV, §7 has been changed to ''Mason-Stothers theorem''.

[**1s**|*4s*] P.xiii, page number for §IX.1: This should be 378.

[**5a**|*6a*] P.xv, last 3 lines: In the first three printings, these three page numbers should be increased by 2. Starting with the third printing, there is a Bibliography beginning on p.895, which should be shown here, and the page number for the Index should be 901. (The Bibliography is expanded in printing 1s, raising the page number of the index to 903 there.)

[**4s**] P.5 line 12 from bottom: ''thereafter'' should be ''henceforth''.

[**3a**|*4a*] P.7, 4/5 of the way down the page, in the line before the italicized statement, after ''respectively'' the line should end ''under suitable conditions. Namely''.

[**1s**|*4s*] P.9, line 5: For ''We define the product'' read ''We define the law of composition''.

[**3a**|*4a*] P.9, third-from-last display: change all three n's to r's.

[**3a**|*4a*] P.12, line 4, change ''which commute'' to ''whose elements commute''.

[**3a**|*4a*] P.15 line 5: $T_{0,\,b}$ should be $T_{1,\,b}$.

[**1s**|*4s*] P.15, line before final display: After ''for example'' add ''letting 0 be the trivial group''.

[**3a**|*4a*] P.17, after first box, ''(1)'' and ''(2)'' should be ''(i)'' and ''(ii)''.

[**1s**|*4s*] P.18: In the last line of Proposition I.3.1, delete comma before ''whose''. At the end of the second line of the proof of that proposition, after ''a cyclic tower'' add ''ending with $\{e\}$''. In the next-to-last line of the page, change ''Theorem 6.4'' to ''Theorem 6.5''.

[**1a**|*2a*] P.19.: Three lines before first (small) display, ''dagonal'' should be ''diagonal''. In the line immediately preceding big matrix: ''$\geq g$'' should be ''≥ 2''.

[**5a**|*6a*] P.19: Two lines after the big matrix, k^r should be k^{n-r}, and on the line after that, ''r-tuple $(a_{1r}, \dots, a_{rn}) \in k^r$. This r-tuple '' should be ''$n-r$-tuple $(a_{1\,r+1}, \dots, a_{n-r,\,n}) \in k^{n-r}$. This $n-r$-tuple''. Also, in printings before the 4th, in the big matrix, the term labeled a_{2r} should be labeled $a_{2,\,r+1}$, the term labeled a_{rn} should be labeled $a_{n-r+1,\,n}$, and between the last two rows of the matrix there should be a ''…'' to show that there may be more (or fewer) than two rows of 0's.

[**6a**|*1s*] P.19: Two lines after the big matrix, k^{u-r} should be k^{n-r}.

[**4s**] P.20, second paragraph: this definition of ''simple'' is repeated in the sentence preceding Theorem I.3.5 on p.22.

[**4a**|*5a*] P.20, first **Example**, third from last line: The condition that in earlier printings appeared as ''r, s relatively prime'' and in the 4th printing as ''r, s relatively prime > 1'' should just be ''r, $s > 1$''.

[**6a**|*1s*] P.23, top: Another reference has been added: R. SOLOMON, A brief history of the classification of the finite simple groups, *Bull. AMS* **38**, 3 (2001) pp. 315-352.

[**5a**|*6a*] P.25, line preceding (**v**): expand ''so'' to ''$\mathbf{Z}/m\mathbf{Z} \cong G/H$, so $n = md$, $m = n/d$ and''.

[**1a**|*2a*] P.26: on line 4, $x(yx)$ should be $x(ys)$. On line 5: $= e$ should be $= s$. On the line after ''**1. Conjugation**'': yxy^{-1} should be xyx^{-1}.

[**1s**|*4s*] p.26, line 7: ''is permutation'' should be ''is a permutation''.

[**5a**|*6a*] P.26, end of the paragraph ''**1. Conjugation**'': Add ''Automorphisms of G of the form γ_x are called **inner**.''

[**5a**|*6a*] P.27, line 3: before ''not a subgroup'' add ''in general''.

[**5a**|*6a*] P.28, next to last line: before ''connected'' add ''where X is''. Also (in all printings including the 6th), on the preceding line I recommend dropping the word ''universal''.

[**3a**|*4a*] P.29 third display should be

$$S = \bigcup_{i \in I} Gs_i \quad \text{(disjoint)}, \quad \text{also denoted } S = \coprod_{i \in I} Gs_i,$$

[**6a**|*1s*] P.30, first line of last display: On the right-hand side of the ''='' sign, insert parentheses around the symbols $\pi(\sigma) f$.

[**3a**|*4a*] P.32: In line 5, after the second word, h_v should be H_v. In the fourth line of ''**Remark ...**'', ''Theorem 5.4'' should be ''Proposition 5.3''.

[**1a**|*2a*] P.36, statement of Proposition I.6.8 ''*porimes*'' should be ''*primes*''.

[**3a**|*4a*] P.36, line before first display: ''be'' should be ''by''.

[**5a**|*6a*] P.36, second line of proof of first **Example**: ''the 5-Sylow subgroup'' should be ''a 5-Sylow subgroup''.

[**3a**|*4a*] P.37, statement of Proposition I.7.1: In the first and second printings, after A add $= \oplus A_i$. In the third printing, after the first sentence add ''Let $A = \oplus A_i$''.

[**1a**|*3a*] P.39, two lines above the **Example**, ''element'' should be ''elements''; and on the same page, two lines above the triangular diagram, M should be $K(M)$.

[**1s**|*4s*] P.39, 5th line after first diagram: For ''is a factor group'' read ''is isomorphic to a factor group''.

[**5a**|*6a*] P.40: Move the line ''The universal group $K(M)$ is called the **Grothendieck group**.'' from two thirds of the way down the page to right after line 7.

[**2a**|*4a*] P.44, line 2: First and second printings: add ''We note that the period of $\bar{b}$ is $\leq$ the period of b. If $n = 0$ we are done. Otherwise''. Third printing: similarly, except that the first of these sentences is already there.

[**5a**|*6a*] P.44, line 7: Change ''The'' to ''Then'', and ''remark'' to ''remarks''.

[**1s**|*4s*] P.45, third line from bottom: $y \in A$ should be $y \in S$.

[**5a**|*6a*] P.48, 4 lines after second display: Change $< n$ to $\leq n$.

[**5a**|*6a*] P.49, next-to-last line: change ''sequence'' to ''sequences''.

[**3a**|*4a*] P.49, second and third lines of statement of Theorem I.9.2: ''*right and left*'' should be ''*left and right*''

[**1s**|*4s*] P.50, first two displays: lim should be $\varprojlim$ in both places.

[**6a**|*1s*] P.51, last paragraph: In the current printing this is formulated in terms of Galois groups, like the preceding paragraph.

[**1s**|*4s*] P.52: On line 5, ''for all n'' should be ''for all r''. In the last line of the proof of Theorem I.10.1, ''direct limit'' should be ''inverse limit''. In the sixth-from-last line of the page, ''countably generated'' should be ''finitely generated''. And on the fourth-from-last line, ''**cofinal**'' should be ''**cofinal** in $\mathcal{F}$''.

[**1s**|*4s*] P.53, statement of **CAT 2**: ''left and right identity'' should be ''right and left identity'' (to match the use of ''respectively'').

[**1s**|*4s*] P.55, next-to-last line: $u_0\, u_0^{-1}$ should be $u \circ u_0^{-1}$.

[**2a**|*3a*] P.56, triangular display: the second ρ should be ρ'.

[**3a**|*4a*] P.56, end of next-to-last line: $x \in G$ should be $x, y \in G$.

[**1s**|*4s*] P.58, 6 lines above ''**Products and coproducts**'': ''situtaions'' should be ''situations''.

[**1s**|*4s*] P.60, line 3: After ''*coproducts exist*'' add '', i.e., every family of objects has a coproduct''.

[**5a**|*6a*] P.60, two lines above first display: Delete ''the complement of x' in'', and if you have a printing before the 4th also delete ''the union of S and''.

[**2a**|*3a*] P.60, three lines above final **Example**: $A \amalg B$ should be $A \amalg B$.

[**1s**|*4s*] P.60, line above example at bottom of page: After ''unique isomorphism'' add ''See the comment, the top of p.58.''

[**5a**|*6a*] P.61, second line after the triangular displayed diagram: Change ''morphisms on Z'' to ''morphisms to X and Y over Z''. If you have a printing before the 4th, then on the preceding line also shorten **direct product** to **product**.

[**1a**|*2a*] P.62, first line and first display: arrow to T should go the other way.

[**1s**|*4s*] P.62, **Example** just before the heading **Functors**: For S read $\mathbb{S}$.

[**5a**|*6a*] P.62, bottom line: change this to ''essentially the same definition, but reversing all arrows $F(f)$, i.e.,''.

[**2a**|*3a*] P.64, line 4: the second $=$ should be $\neq$.

[**5a**|*6a*] P.67, equation in the line after the triangular display: g_* should be ψ_*.

[**3a**|*4a*] P.71, middle of page: the displayed equation $f_i \circ g_* = g_i$ should be $g_* \circ f_i = g_i$.

[**1s**|*4s*] P.71, last line of text: For ''This is done by representing the group as follows'' read ''The group can be represented as follows''.

[**5a**|*6a*] P.74, first line of Corollary V.12.8: After ''*groups*'' add ''*with* $G_i \cap G_j = \{1\}$ *if* $i \neq j$''.

[**1s**|*4s*] P.76, first display: The right-hand side should be $(x_1 \psi(h_1) x_2,\ h_1 h_2)$.

[**2a**|*3a*] P.76, Exercise 16: ''Chapter XVIII'' should be ''Chapter XIII''.

[**1a**|*2a*] P.77, Exercise 22, first line: ''grop'' should be ''group''.

[**5a**|*6a*] P.79: Exercise 38(d) should begin ''Assume that n is prime.''

[**1s**|*4s*] P.81, Exercise 51(b): In the next to last line, in the phrase ''into fiber products in the category of sets'', delete the word ''fiber''. (That is, the functor turns fiber products into products.)

[**6a**|*1s*] P.82: Exercise 57 has been added, concerning group actions, with topological applications. The next correction applies to the version of that added exercise in [**1s**]:

[**1s**|*4s*] P.82, next-to-last line: Change ''closed. Then'' to ''closed, and that''.

[**1s**|*4s*] P.84, next-to-last display: The $x_i x_j$ at the end should be $x_i y_j$.

[**3a**|*4a*] P.85, line 4: $x \in A$ should be $x \in S$.

[**1s**|*4s*] P.92, near top, second line of **Examples**: For ''more than 2'' read ''at least 2''.

[**5a**|*6a*] P.93, definition of **field**: Before ''ring'' add ''commutative''. (Also, at the beginning of that line he has ''defined'' for ''define'', even in the 6th printing.)

[**4a**|*5a*] P.99, 4th and 3rd lines from bottom, Delete ''$f \in A[X]$ *and*''.

[**1s**|*4s*] P.100: In the line after the Example, the summation should start with $i = 0$. In the third from last line of the page, ''polynomial'' should be ''polynomials''.

[**3a**|*4a*] P.102, immediately after third display: change ''is an isomorphism'' to ''is injective''.

[**3a**|*4a*] P.105, first line, left-hand side: change (t) to (z).

[**3a**|*4a*] P.106, Proposition II.3.1: add at the end the condition ''*and* $h(a) = a$ *for all* $a \in A$''.

[**3a**|*4a*] P.107, statement of Proposition II.3.2, last line (displayed equation): on the right-hand-side, the h should be f. (The h on the left-hand side is correct.)

[**1s**|*4s*] P.107, computation in middle of page: The last line should be $= h(\alpha)h(\beta)$.

[**4a**|*5a*] P.107, display beginning $h(\alpha\beta)$: The second line of this display should be a double summation $\Sigma\Sigma\, f(a_x)\, f(b_y)z$, with the respective ranges of summation as on the preceding line.

[**1s**|*4s*] P.112, second line: change (a, b) to $(a) + (b)$.

[**4s**] P.112, fifth line: same correction as above.

[**6a**|*1s*] P.112, line after the display ''$(a_n) \subseteq (a) \subseteq (a_n)$'': The conclusion after ''Hence'' does not quite follow from what has been said, but becomes correct if one assumes that in the construction of the chain of (a_i), one does not stop at a finite step (a_n) unless the ideal at that step is maximal in S. In printing 1s, Lang has changed this part of the proof, using the preceding sentence to deduce that S is inductive, and applying Zorn's Lemma. My notes for this this page give the approach I prefer.

[**1s**|*4s*] P.112, first line of paragraph starting just below middle of page: ''a_n'' should be ''a''.

[*6a*|**1s**|*4s*] P.113, second line before first Example: **relaively** should be **relatively**.

[**1s**|*4s*] P.115: In the next-to-last line of Exercise 5, change ''those primes of p'' to ''of the form up where p is a prime'', and at the end of the last line add the phrase ''and u is a unit of $S^{-1}A$''. In the first line of Exercise 10: ''all element'' should be ''all elements''.

[**1s**|*4s*] P.116, last line: add ''; Chapter III, Exercises 11-13.''

[**4s**] P.118, fifth line below the word **Examples**, for ''two-sided'' read ''left''. Three lines later, ''More general'' should be ''More generally''.

[**1s**|*4s*] P.118, second line of the paragraph after the **Examples**: ''$a + J$'' should be ''$x + J$''.

[**1s**|*4s*] P.126, line 6 of **Examples: ...** : $F(K') \to$ should be $F(K) \to$.

[**1a**|*3a*] P.128, three lines above the proposition: ''9.1'' should be ''7.1''.

[**6a**|*1s*] P.136, two lines above boxed display: ''left ideal'' should be ''two-sided ideal''.

[**4a**|*5a*] P.142, first two sentences of §III.6, Delete the word ''finite'' in the first sentence. In the second sentence, for ''1-dimensional space'' read ''free module of rank 1''.

[**1a**|*3a*] P.143, line below that on which **dual basis** is defined: ''B of E'' should be ''B of $E^{\vee}$''.

[**6a**|*1s*] P.145, line 3: **P4** should be **P2**.

[**6a**|*1s*] P.154, Theorem III.7.9: Add an initial sentence, ''*Assume that the elementary matrices in* R *generate* $GL_n(R)$''. In printing 1s Lang has also added at the end of the section a reference to a paper showing that this condition does not always hold.

[**6a**|*1s*] P.155, last line: T should be Γ.

[**6a**|*1s*] P.159: In the line after the first display, ''on cokernels'' should be ''on kernels''. Two lines after the third display, ''image of g'' should be ''image of g_*''.

[**6a**|*1s*] P.160, bottom line: $f^i_j(x)$ should be $-f^i_j(x)$.

[**6a**|*1s*] P.161, three lines above second display: ''homorphism'' should be ''homomorphism''.

[**1s**|*4s*] P.163, third line of **Examples**: $k[T]$ should be $k[[T]]$.

[**1s**|*4s*] P.165, Exercise 2: Before ''commutative ring'' insert ''non zero''.

[**1s**|*4s*] P.169, Exercise 17(a): Change ''a projective system'' to ''an inverse system''.

[**1s**|*4s*] P.170, Exercise 17(b): Change ''all ideals'' to ''all non zero ideals''.

[**6a**|*1s*] P.172: Three new exercises, on graded algebras, have been added on this (previously blank) page.

The next comment applies to the version of these in printing [**1s**].

[**1s**|*4s*] P.172, line after top display: Delete the sentence ''In particular, A is an A_0-algebra''. (So the word ''algebra'' in later sentences should be understood as meaning ''k-algebra''.)

[**3a**|*4a*] P.175, line after first display: $f - gg$ should be $f - qg$.

[**2a**|*3a*] P.175, five lines above Theorem IV.1.4, add ''A polynomial is called **monic** if it has leading coefficient 1.'' Add to the index a reference to this page for the term ''monic polynomial''.

[**1s**|*4s*] P.176, first display, and line before third display: f_i should be f_j in both places.

[**3a**|*4a*] P.178, line 6 from bottom: ''a polynomial'' should be ''a nonzero polynomial''.

[**1s**|*4s*] P.179, display in statement of Proposition IV.1.12, and the following display: The summations should start with $\nu = 0$ and $\mu = 0$ respectively.

[**5a**|*6a*] P.181, Theorem IV.2.1, third line of *Proof*, end of line: Change $= 1$ to $= 0$.

[**3a**|*4a*] P.182, statement of Theorem IV.2.3: change ''*are either primes of* A *or ...*'' to ''*are the primes of* A *and ...*''.

[**5a**|*6a*] P.182, 6 lines from bottom: move the phrase ''From the Gauss lemma'' to modify ''c = cont(f)'' on the preceding line.

[**3a**|*4a*] P.184, end of line 10: delete ''$r \neq m$ and''.

[**2a**|*3a*] P.184, line 3 from bottom: ''E's'' should be ''Eisenstein's''.

[**3a**|*4a*] P.186, Theorem IV.4.1, first line of Proof: The second sentence should begin, ''Let $\mathbf{a}_i$ consist of 0 and the set of elements $a \in A$ ...'' (but with $\mathbf{a}_i$ having the $\mathbf{a}$ in fraktur font, as in the middle of the line after the display).

[**1s**|*4s*] P.186, first line of proof of Theorem IV.4.1: In ''a_i'', the ''a'' should be fraktur, as in the line after the following display, and $A[X]$ should be $A[X]$.

[**6a**|*1s*] P.187, three lines before Corollary IV.4.2: ''numer'' should be ''number''.

[**5a**|*6a*] P.187, statement of Corollary IV.4.2: Change ''*a finitely generated commutative ring containing* A'' to ''*a commutative ring finitely generated over* A''.

[**3a**|*4a*] P.190, from line preceding last display to bottom of page: everywhere change the notation f^σ to σf. In particular, on the line following the display, change the incorrect formula $f^{\sigma\tau} = (f^\sigma)^\tau$ to the correct formula $\sigma\tau f = \sigma(\tau f)$.

[**1s**|*4s*] P.191, statement of Theorem IV.6.1, second line: Change ''$\leq d$'' to ''d''.

[**3a**|*4a*] P.193, 6th line of **Cubic Case**: $X_2^n X_3^m$ should be $X_2^m X_3^n$.

[**1s**|*4s*] P.193: In the line after the third-from-last display, ''$D = -va^3 = -v$'' should be ''$D = va^3 = -v$'' (i.e., Lang's first minus sign is wrong, while the second is right). And in the line after the next-to-last display, $= 2b^2 = w$ should be $= wb^2 = w$.

[**5a**|*6a*] P.194, 2nd and 4th displays: Reverse the signs on all four ''$\frac{1}{3}$''s (three in the 2nd display and one in the 4th).

[**6a**|*1s*] P.200, end of §IV.8: A reference has been added to the material noted under p.220 below.

[**1s**|*4s*] P.210, first line: ''*post*'' should be ''*posto*''.

[**1s**|*4s*] P.213, first line of Exercise 7: Change ''q elements'' to ''$q = p^m$ elements''.

[**1s**|*4s*] P.214, Exercise 9: After ''automorphism of $A[X]$'' add the words ''inducing the identity on A''.

[**3a**|*4a*] P.214, Exercise 10: Change the first sentence to ''Let K be a field and $K(X)$ the quotient field of $K[X]$.''

[**1s**|*4s*] P.216, first display: b_{d-1} should be b_{d-2}.

[**3a**|*4a*] P.217, line 5, end of line: ''in (a)'' should be ''in (b)''.

[**6a**|*1s*] P.220: Lang has added a paragraph and four references, indicating why what he previously called ''Mason's Theorem'' is now called the ''Mason-Stothers theorem'', and discussing further proofs thereof.

[**3a**|*4a*] P.224, statement of Proposition V.1.2, line 3: ''*a basic*'' should be ''*a basis*''.

[**3a**|*4a*] P.225, middle of page, line following display: ''is **finite**'' should be ''is called **finite**''.

[**5a**|*6a*] P.230, next-to-last line of *Proof* near top of page: before $[E' : k]$ delete ''as''. In second and

third lines of next paragraph, change ''Then EF is the quotient field of $E[F]$, and also the quotient field of $F[E]$'', to ''Then $E[F] = F[E]$ and EF is the quotient field of this ring''.

[**1s**|*4s*] P.233, statement of Proposition V.2.7: change ''$\leq$ *the number of roots of* p'' to ''$\leq \deg p$'', and add at the end of the second line the words ''*in* k^{a}''.

[**6a**|*1s*] P.235, third line of **Examples**: after ''other automorphisms'' add ''$\neq$ id''.

[*6a*|**1s**] P.235, third line of **Examples**: after ''$\neq$ id'', there should not be a ''.''.

[**3a**|*4a*] P.236, Theorem V.3.1, last line. Change ''*on* K'' to ''*onto* K''. The first two printings also have ''*including*'' instead of ''*inducing*'' on the third line of the theorem.

[**3a**|*4a*] P.239, diagram in middle of page: add a vertical line between F and E. In the first printing (only), λ also goes in the wrong direction: it should go from right to left.

[**3a**|*4a*] P.241, line 3: change ''if α is separable over k, and F is any extension of k'' to ''if $k \subset F \subset K$ and $\alpha \in K$ is separable over k''.

[**3a**|*4a*] P.242, line 5: ''Theorem 4.2'' should be ''Corollary 4.2''.

[**4s**] P.243, line below first display: ''the same field'' should be ''the above field''.

[**3a**|*4a*] P.243, two lines above last display: after ''roots of f'' add ''in a fixed algebraic closure''.

[**3a**|*4a*] P.247, Theorem V.5.5, first line: change ''*Then* $\mathbf{F}_{p^n}$'' to ''*Then in any algebraic closure of* $\mathbf{F}_p$, *the subfield* $\mathbf{F}_{p^n}$''.

[**6a**|*1s*] P.247, last line of Theorem V.5.5: After *cyclic* add *of order d*.

[**3a**|*4a*] P.248, first display: change the three occurrences of the index i to j (since i was given a particular meaning on the preceding page).

[**3a**|*4a*] P.249, Proof of Corollary V.6.4: change ''Obvious'' to ''Immediate by Theorem 4.1''.

[**4s**] P.249, statement of **P.Ins. 2**: For ''the irreducible equation ... is $X^{p^n} - a = 0$'' read ''the irreducible polynomial ... is $X^{p^n} - a$''.

[**3a**|*4a*] P.251: On line 2, K should be k. In the fifth line of the proof of Corollary V.6.10, the phrases ''purely inseparable'' and ''separable'' (in two successive sentences) should be interchanged. In the next line, change ''Iterating'' to ''Similarly''.

[**6a**|*1s*] P.251, third line of statement of Proposition V.6.11: K_G should be K^G.

[**6a**|*1s*] P.252, line after first display: σ_1 should be σ_i.

[**3a**|*4a*] P.252, Corollary V.6.12: Connect the two sentences with ''and'' (so that the hypothesis applies to both).

[**1a**|*2a*] P.256, Exercise 26, second line: before ''irreducible'' add ''monic''.

[**5a**|*6a*] P.256, Exercise 27, second display: X_{n-1} should be X_{n+1}.

[**3a**|*4a*] P.262, line 7: Here Lang indicates the notation $G(K/k)$ or G for the Galois group, and he will indeed use these for the next few sections. But eventually, other notations start cropping up as well; so after ''$G(K/k)$'', add the alternatives ''$G_{K/k}$, $\mathrm{Gal}(K/k)$'', as Lang has done in later printings. Also, on line 7 from the bottom, change ''Chapter VII'' to ''Chapter V''.

[**5a**|*6a*] P.263, two lines above Lemma VI.1.7: Before ''Corollary 1.6'' add ''the last assertion of''.

[**6a**|*1s*] P.264, next-to-last line of proof of Theorem VI.1.8: ''oder'' should be ''order''.

[**3a**|*4a*] P.264, two lines before Corollary VI.1.9: again, ''Chapter VII'' should be ''Chapter V''.

[**6a**|*1s*] P.266, proof of Corollary VI.1.1: ''... a subgroup of an abelian group is abelian'' should say ''... is normal''.

[**1s**|*4s*] P.269, one line after the proof of Theorem 1.17: ''composite'' should be ''compositum''.

[**6a**|*1s*] P.269, first line of §VI.2: Before ''polynomial'' add ''separable''.

[**3a**|*4a*] P.270: On line 8, ''and assume f is separable'' should be ''Since $\mathrm{char}\, k \neq 2, 3$, f is separable''. On the 6th line from the bottom, ''order 2'' should be ''degree 2''

[**3a**|*4a*] P.271: In line 2 (display), Q should be $\mathbf{Q}$. Also, two lines above the diamond-shaped diagram, ''sending i on $-i$'' should be ''sending i to $-i$'', and similarly four lines below the same diagram ''on the root'' should be ''to the root''.

[**1s**|*4s*] P.274, second line before Example 8: ''[123$\cdots p$], which generate S_p'' should be ''[12345], which generate S_5''.

[**1s**|*4s*] P.275, lines 5 and 7: P_{n-1} should be P_{n+1} in both places.

[**6a**|*1s*] P.275, paragraph starting below middle of page: After the first sentence, ''Over the field $\mathbf{C}(t)$ ...'', Lang has added a sentence referring to papers on this and a related case. (To make room for this, he has un-displayed two of the four preceding displays, so the page layout looks different; but the content seems the same except for this added sentence.)

[**4a**|*5a*] P.277, three lines above the next-to-last display: In the first three printings, ''prime to'' should be ''not divisible by''. In the 4th printing, ''by to'' should be ''by''.

[*4a*|**5a-1s**|*4s*] P.277, third line from the bottom: ''$G_{\kappa(\zeta)/\kappa}$'' should be $G_{k(\zeta)/k}$.

[**2a**|*3a*] P.277, second line from the bottom: ''$\mathbf{Z}/n\mathbf{Z})^*$'' should begin with an open-parenthesis.

[**1s**|*4s*] P.277, second line from the bottom: $\kappa =$ should be $k =$.

[**3a**|*4a*] P.279, line 5: change ''had seen'' to ''saw''.

[**2a**|*3a*] P.279, display with a fraction-line: the Φ in the denominator should be Φ_d.

[**1a**|*3a*] P.281: on lines 4 and 5, $n \geq e$ should be $j \to \infty$, and ''Conrad'' should be ''Bang''. Two lines later, after ''the coefficient of X^7,'' add ''and X^{41},''.

[**3a**|*4a*] P.281: in the line before the central display with braces, after ''a nonzero integer'' add ''not divisible by p'', and in that display itself, add to the first line, ''for some x'', and to the second line, ''for all x''.

[**4s**] P.282: The extra-long minus sign appearing twice in the third display is a typographical error, with no special meaning.

[**1a**|*2a*] P.283: In the 4th line of the second paragraph of **Examples**, ''Similary'' should be ''Similarly''. And in the line before Theorem VI.4.1: ''thsese'' should be ''these''.

[**3a**|*4a*] P.284, last two displays: As with the notation for Galois groups (first erratum re p.262 above), Lang will use other notations along with those given here. So change ''$N_k^E(\alpha) =$'' to ''$N_{E/k}(\alpha) = N_k^E(\alpha) =$'', and ''$\mathrm{Tr}_k^E(\alpha) =$'' to ''$\mathrm{Tr}_{E/k}(\alpha) = \mathrm{Tr}_k^E(\alpha) =$'', as Lang does in later printings.

[*3a*|**4a-5a**|*6a*] P.284, last display: $T_{r_{E/k}}$ should be $\mathrm{Tr}_{E/k}$.

[**3a**|*4a*] P.285: In the second line of the second paragraph of the Proof, change ''an embedding of E in k^{a}'' to ''an automorphism of k^{a}''. And on the 4th line from bottom, change ''penultimate'' to ''next to highest''.

[**1s**|*4s*] P.286, two and six lines before Corollary VI.5.3: $\hat{E}$ should be $E^\vee$. (The earlier, but not the later of these errata was corrected in printing [*6a*].)

[**1s**|*4s*] P.286, statement of Corollary VI.5.4: change ''*the distinct set of embeddings*'' to ''*the distinct embeddings*''.

[**1s**|*4s*] P.288, first and second lines after the last display in the proof of Proposition VI.5.6: $N_{F/k}(\alpha)^r$ and $N_{F/k}(\alpha)^d$ should be $N_{F/k}(\alpha)$ and $N_{F/k}(\alpha)^r$ respectively.

[**1s**|*4s*] P.289: On the second line before Theorem VI.6.2, ''when we apply σ'' should be ''when we apply $\beta\sigma$''. And on the second line of that theorem, after ''*of* k'' add ''(*if not* 0)''. (I suggested that Lang replace ''*prime to*'' on the preceding line by ''*not divisible by*'', but this was the change he made instead. It has the same effect.)

[**3a**|*4a*] P.289, line 4 from bottom: G should be G.

[**3a**|*4a*] P.290, Theorem VI.6.4, third line of Proof: K should be K.

[**4s**] P.290, next-to-last line: $1, \ldots, p$ should be $1, \ldots, p-1$.

[**3a**|*4a*] P.291, line 6: after ''[co]efficient of X^{d-1},'' add ''in g''.

[**1s**|*4s*] P.293, third line before **Remark**: Before ''prime degree'' add ''of''.

[**5a**|*6a*] P.293, 5th line of Remark: ''are'' should be ''is''.

[**1s**|*4s*] P.294, second line on page, and also first line of statement of Theorem VI.8.1: Again, after ''characteristic of k'' insert ''(if not 0)''. (And again, changing ''prime to'' to ''not divisible by'' would have the same effect.)

[**6a**|*1s*] P.294, line just above the next-to-last display on the page: Delete the final word, ''hence''.

[**5a**|*6a*] P.297, 8 lines from bottom: After ''hence'' add ''inductively''.

[**1s**|*4s*] P.297, last line: α should be a.

[**5a**|*6a*] P.298, 2nd display: The index v (occurring twice) should be v, as in the preceding display. (Also, if you have a printing before the 4th, the value on top of the Π should be p, as in the preceding display.) On the same page, 8 lines from the bottom, $\pm c^2 i$ should be $\pm 2c^2 i$.

[*3a*|**4a-5a**|*6a*] P.299, lines 8 and 15 from bottom: k^a should be k^{a}.

[**1s**|*4s*] P.299, last display and line following it: v should be v in both places.

[**5a**|*6a*] P.300, 2 lines before Theorem VI.9.4: ''given'' should be ''gives''. Also, if you have a printing before the 4th, in the second display on that page, the exponent $b(n)$ should be $b(\sigma)$.

[**4s**] P.300, 4th line of proof of Theorem VI.9.4: ''given'' should be ''give''.

[**3a**|*4a*] P.301, line 9 from bottom: ''had never published'' should be ''never published''.

[**1s**|*4s*] P.303, end of second line of the proof of Lemma VI.10.2: Add a second '')''.

[**5a**|*6a*] p.321, end of second line of Exercise 5: ''of f'' should be ''of F''.

[**1s**|*4s*] P.322: On line 5, $X^4 + 4x^2 + 2$ should be $X^4 + 4X^2 + 2$. On the next-to-last line of Exercise 8(b), after ''to a subgroup'', add ''of''. On the second line of Exercise 11, change ''a real subfield k of the complex numbers'' to ''a subfield $k \subset \mathbf{R} \subset \mathbf{C}$''.

[**3a**|*4a*] p.322 Exercise 8, part (b), line after display: the case $\alpha^2 - \beta^2 \in \mathbf{Q}$ does not actually occur. It corresponds to a subgroup of $D_8 \subseteq S_4$ which is indeed isomorphic to $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, but which is not transitive on $\{1,2,3,4\}$, so it is excluded by the assumption that f is irreducible.

[*3a*|**4a-5a**|*6a*] p.322 Exercise 8, part (b), last two lines: D_∞ should be D_8, and in $Z/2Z \times Z/2Z$, the Z's should be $\mathbf{Z}$'s, and the sentence should (and in the 5th printing does) end with a close-parenthesis.

[**5a**|*6a*] P.322, bottom or p.323 top: Add the two lines

> contains enough cycles to generate S_n. Use the Chinese remainder theorem, also to be able to apply Eisenstein's criterion.]

[**1s**|*4s*] P.323, ''Note'' starting after Exercise 20: Delete the ''['' at the beginning of the Note, which is never closed.

[**1s**|*4s*] P.325: On the top line, after ''and G'' insert ''$\neq \{1\}$''. On the third line: After ''non-zero'' insert ''relatively prime''.

[**3a**|*4a*] P.327: On the last line of Exercise 38, at the beginning of the line add ''or $2d_n$''. On the 5th line of Exercise 39, K should be K (3 occurrences).

[**1s**] P.356, statement of Theorem VIII.1.1: There is a space in the last line of the statement of the theorem, which should contain a $\mathfrak{B}$.

[**5a**|*6a*] P.356, statement and proof of Theorem VIII.1.1. The proof given is incorrect. Make the following changes:

In the statement of the theorem, change ''*If* Γ *is a set of generators of* K *over* k'' to ''*If* Γ *is a subset of* K *such that* K *is algebraic over* $k(\Gamma)$''.

In the second line of the proof, after $m \geq 1$ add '', m minimal''. Three lines later, before ''polynomial'' add ''irreducible''.

Now, for everything from the word ''Furthermore'' after the first display through the last display on the page, substitute the following. (What follows is approximately the corrected proof that Lang has, starting with the 6th Addison-Wesley printing, but incorporates the main clarifications to that proof that I give in the body of this Companion.)

> This is true by the sentence preceding the theorem, and the fact that we can clear denominators to turn a polynomial with coefficients in $k(x_1, \ldots, x_m)$ into one with coefficients in $k[x_1, \ldots, x_m]$. Now since $w_1, \ldots, w_n$ are algebraically independent, f_1 must involve at least one x_i; by renumbering $x_1, \ldots, x_m$ we may assume this is x_1, and

so write $f_1 = \Sigma\, g_j(w_1, x_2, \dots, x_m)x_1^j$ with $g_N \neq 0$ for some $N \geq 1$. No irreducible factor of g_N vanishes on $(w_1, x_2, \dots, x_m)$, otherwise w_1 would be a root of two distinct irreducible polynomials over $k(x_1, \dots, x_m)$ (namely f_1 and this irreducible factor, which by the assumption that it vanishes at the indicated point must involve w_1 by algebraic independence of the x's, and which cannot coincide with f_1 because it does not involve x_1). Hence f_1 is a nonzero polynomial satisfied by x_1 over $k(w_1, x_2, \dots, x_m)$, so x_1 is algebraic over that field, hence $w_1, x_2, \dots, x_m$ are algebraically independent over k, otherwise the minimality of m would be contradicted.

Suppose inductively that after a suitable renumbering of $x_2, \dots, x_m$ we have found that K is algebraic over $k(w_1, \dots, w_r, x_{r+1}, \dots, x_m)$. Then there exists a nonzero polynomial f in $m+1$ variables with coefficients in k such that

$$f(w_{r+1}, w_1, \dots, w_r, x_{r+1}, \dots, x_m) = 0.$$

Since the w's are algebraically independent over k, it follows by the same argument as in the first step that some x_j, say x_{r+1}, is algebraic over $k(w_1, \dots, w_{r+1}, x_{r+2}, \dots, x_m)$.

Now continue in Lang, from ''Since a tower ...'', and, after finishing those four lines, go to the comment on the last line of that page in the body of this Companion.

[**6a**|*1s*] P.364, 10 lines from bottom: ''Proposition (iii)'' should be ''Proposition 3.3''.

[**4s**] P.364, next-to-last line: the name **MacLane** should be **Mac Lane** (with a space before the L). The same correction applies to the reference to Mac Lane in the index.

[**6a**|*1s*] P.366: In the first line of Lemma VIII.4.10, ''alebraically'' should be ''algebraically''. In the sixth line from bottom, ''Lemma 4.11'' should be ''Lemma 4.10''.

[**1s**|*4s*] P.375, Exercise 6(a): $k(y_1, \dots, y_r)$ should be $k_u(y_1, \dots, y_r)$.

[**1s**|*4s*] P.425, Proof of Theorem X.4.4: In the display, ''E'' should be F, the second $\subset$ should be $\oplus$, and the final ''$\supseteq$'' should be $\oplus$. Also, throughout the proof, P_o is a typo for P_0 (but the notation is at least consistent). If you have an Addison Wesley printing, see next correction instead.

[**6a**|*1s*] P.425, Theorem X.4.4 (but if you have a Springer printing, see preceding correction instead): In the first line of the statement, before *projective* add *finite*.

The proof given assumes incorrectly (in the 3rd and 4th lines) that A has no zero-divisors; so use instead the following proof:

Given $x_1, \dots, x_n \in E$ as in the sentence beginning ''In fact'', let F be a free module with basis $e_1, \dots, e_n$, and $f: F \to E$ the homomorphism taking each e_i to x_i. We want to prove f an isomorphism. By Lemma 4.3, f is surjective. Since E is projective, it follows that f splits, i.e., we can write $F = P_0 \oplus P_1$, where $P_0 = \mathrm{Ker}(f)$ and P_1 is mapped isomorphically to E by f. Now the linear independence of $x_1, \dots, x_n \bmod \mathbf{m}E$ shows that

$$P_0 \subseteq \mathbf{m}F = \mathbf{m}P_0 \oplus \mathbf{m}P_1,$$

hence $P_0 \subseteq \mathbf{m}P_0$. Also, as a direct summand in a finitely generated module, P_0 is finitely generated. So by Lemma 4.3, $P_0 = (0)$, and f is an isomorphism, as was to be proved.

The final statement follows because we can complete a sequence of the indicated sort to one that is linearly independent over $A/\mathbf{m}$, and apply the first result.

[**1s**|*4s*] P.426, 2nd line: ''lineary'' should be ''linearly''.

[**6a**|*1s*] P.468, fourth from last line of §XII.1: ''*u*t'' should be ''to''.

[**6a**|*1s*] P.470, top line: Change ''$\leq$'' to ''$=$''.

[**5a**|*6a*] P.470, next-to-last display: The ξ or χ (depending on which printing you have) should be an x.

[**5a**|*6a*] P.472, various displays: All + signs should be − signs. (In recent printings there were just

three such signs still needing correcting, in the 1st, 5th and 6th displays. In the first one or two printings, there were more.)

[**4s**] P.476 reference [Se 62]: ''complêtement'' should be ''complètement''. (In printings through [**1s**], the same error also occurs in this reference on p.899 or p.900, depending on printing.)

[**6a**|*1s*] P.512: At end of line, add ''with $j = i+1$''.

[**1a**|*3a*] P.519, third line from bottom: ''Proposition 4.15'' should be ''Proposition 4.16''.

[**6a**|*1s*] P.520, Proposition XIII.4.19: Change ''*Let* Δ *be*'' to ''*Let* $\{\Delta\}$ *be*''.

[**6a**|*1s*] P.522, line after last display: ''subscript'' should be ''prefix''.

[**6a**|*1s*] P.529, first two lines of proof of Proposition XIII.6.1: ''Fix a basis of E'' should be ''Fix bases of E and F'', and ''elements of E'' should be ''elements of these modules''.

[**1s**|*4s*] P.531, second display: The first ''='' sign should be a ''+''.

[**6a**|*1s*] P.547, Exercise 25: Add the condition $n > 1$.

[**4s**] P.549, parenthetical sentence before list of references: this should end with '')''.

[**6a**|*1s*] P.550, exercise 32: Lang originally cited only a paper of his with Kolchin. As of the first Springer addition, he also shows the much older reference, A. Speiser, Zahlentheoretische Sätze aus der Gruppentheorie, *Math. Z.* (1919) **5**, pp.1-6.)

[**6a**|*1s*] P.552: There is a new exercise, 36, on irreducibility of $\mathbf{sl}_n(F)$ under conjugation by $SL_n(F)$.

[**6a**|*1s*] P.555, next-to last display: The v's should all be υ's.

[**6a**|*1s*] P.561, second line after second display: R should be k.

[**6a**|*1s*] P.565, two lines above third display: For ''its expansion according to the first column'' read ''its expansion as a sum over permutations''.

[**6a**|*1s*] P.570: There is a new exercise, 26, on the subgroup of $SL_n(\mathbf{C})$ consisting of diagonal positive matrices with positive real entries, and its normalizer.

[**1s**|*4s*] P.603, last line: ''over k'' should be ''over R''.

[**1s**|*4s*] P.611: In the line after the third-from-last display, the comma after ''id'' should be a period. And in the next-to-last line, x'' should be x, and E'' should be E.

[**6a**|*1s*] P.617, next-to-last line: $E_1 \otimes E_2$ should be $E_1 \oplus E_2$.

[**1s**|*4s*] P.618, line following the second diagram: $E_1 \otimes E_2$ should be $E_1 \oplus E_2$.

[**6a**|*1s*] P.618, third-from-last line: ''Proposition 2.6'' should be Proposition 2.7.

[**6a**|*1s*] P.619, ''Remark on abstract nonsense'': In the fifth line, after the assumption that $B \mapsto T(A, B)$ is right exact for each A, add the assumption that $A \mapsto T(A, B)$ is right exact for each B. In the next two lines, ''T-exactness'' is defined for objects of both categories; however, in the case of objects of $\mathcal{B}$ he should call this ${}^tT$-exact (i.e., exact as a second argument of T) since he uses that notation in the next paragraph. In the third paragraph, for ''Instead of an ideal **a** in R'' read ''For the analog of Proposition 3.7''.

[**1s**|*4s*] P.619: In the fourth and fifth lines above the first display, interchange ''**${}^tT$-exactness**'' and ''**T-exactness**''. In the second line above that display, B should be $\mathcal{B}$. In the third line after the second display, the second word, ''then'', should be ''taken''.

[**6a**|*1s*] P.621: In the first line of Lemma XVI.3.10, delete ''*finite*''. And in the last display, $\beta_j \otimes y_j$ should be $\beta_j y_j$.

[**6a**|*1s*] P.623, line preceding Proposition XVI.4.1: ''From Proposition 1.2'' should be ''From Proposition 2.3''.

[**6a**|*1s*] P.628, Corollary XVI.5.4: Change ''*functorial*'' to ''*natural*''.

[**6a**|*1s*] P.633, line 4: ''(mapping k ...)'' should be ''(mapping R ...)''.

[**6a**|*1s*] P.639, Hint to Exercise 13: where Lang speaks of (ii)$\Rightarrow$(iii) holding by ''the preceding exercise'', he means by Exercise 11. In the third-to-last line of the Hint, (iii) should be (iv). In the last sentence, *Algebre* should be *Algèbre*.

[**6a**|*1s*] Pp.639-640: There is a new exercise, 36, on the Casimir element of $E \otimes E$, where E is a vector space (over any field) with a nondegenerate symmetric bilinear form.

[**3a**|*4a*] P.645, first line of Lemma XVII.2.1: ''*nonnecessarily*'' should be ''*not necessarily*''.

[**6a**|*1s*] P.786, next-to-last display: Mark the arrow $M \to T$ as f. (It is referred to by that letter a few lines later.)

[**6a**|*1s*] P.830, Exercise 20: Before ''ring'' add ''Noetherian''.

[**4s**] P.876, line preceding the definition of **finite**: for ''of course'' read ''in particular''. Also, if you have a printing before [*6a*], in the second line of Proposition A2.1.1 change ''*namely*'' to ''*say*''.

[**1s**|*4s*] P.877, second line of proof of Proposition A2.1.4: What looks like x_v is an x_y that didn't print right.

[**3a**|*4a*] P.877, two lines before Proposition A2.1.5: ''and is denumerable'' should be ''and D is denumerable''.

[**1s**|*4s*] P.880: On line 6, ''Example 2'' should be ''Example 1''. And on the fifth line of the last paragraph of the page, the J near the end of the line should be J'.

[**3a**|*4a*] P.880, two lines below Zorn's Lemma: the words ''Chapters I, §10 and XV, §2'' should read ''Chapters I, §7 and XIV, §2''.

[**3a**|*4a*] P.881: On the line following the second display, and three times in next line, change w to z, since w already means something else. And in condition **3.** later on the page, before ''totally ordered'' add ''non-empty''.

[**3a**|*4a*] P.883, proof of Lemma A2.2.3: Starting two lines below the first display, the proof is garbled. It is corrected in later printings; I won't go into details, because I prefer the development in my handout on the Axiom of Choice etc.

[**3a**|*4a*] P.884, Corollary A2.2.4: change the three occurrences of S in the proof (2nd and 3rd lines) to A.

[**5a**|*1s*] P.885: Until the first Springer printing, Lang did not mention comparability of cardinals (see my handout on *The Axiom of Choice, Zorn's Lemma, and all that*), though he used that fact on p.889. Subsequently, he added it to this page – in two different places. Since one suffices, I will indicate the shorter one: Before the italicized line ''*Let* f: ...'' on this page, insert: ''Using Zorn's Lemma, it is easy to show (see Exercise 14) that $\mathrm{card}(A) \leq \mathrm{card}(B)$ or $\mathrm{card}(B) \leq \mathrm{card}(A)$.''

[*6a*|**1s**|*4s*] P.885, first display: ''$\mathrm{card}(A \leq$'' should be ''$\mathrm{card}(A) \leq$''.

[**4a**|*5a*] P.886, first line: $f(b)$ should be $g(b)$.

[**1s**|*4s*] P.886, end of line 4: $H(f^{-1}(b))$ should be $h(f^{-1}(b))$.

[**1s**|*4s*] P.888, second line of proof of Theorem 3.6: ''$f: B \times B$ is a bijection'' should just be ''f is a bijection''.

[**1s**|*1s*] P.889: Delete ''$\mathrm{card}(M) \leq$'' from the first display, and ''by Bernstein's Theorem'' from the next line.

Reading further in the paragraph, you will see that Lang considers the two cases $\mathrm{card}(C) \leq \mathrm{card}(M)$ and $\mathrm{card}(M) \leq \mathrm{card}(C)$, and then, on the very last line (added in the 6th Addison Wesley printing), to be able to assert that one or the other of these must hold, he calls on Exercise A2:L14, which sketches a proof of Comparability of Cardinals. We saw that result proved from Zorn's Lemma in the handout on the Axiom of Choice.

[**5a**|*6a*] P.889, Corollary A2.3.8: After ''*nonempty sets*'' add ''*with* A_n *infinite*''.

[**3a**|*4a*] P.890 line 3: ''Corollary 3.4'' should be ''Corollary 3.7'' here, and again 11 lines from the bottom of the page.

[**1s**|*4s*] P.892, fifth line of proof of Theorem A2.4.1: ''the initial segment'' should be ''an initial segment''.

[**3a**|*4a*] P.892, 4th from last line of Proof: X_j should be $X_j \cap Y$.

[**6a**|*1s*] P.892, 3rd from last line of Proof: ''$(X, \omega$'' should be ''(X, ω)''.

[**4a**|*5a*] P.893, Exercise 4: the comma after the exercise-number should be a period.

[**1a**|*5a*] P.893, Exercise 13. Lang notes in brackets a question whose answer he ''doesn't know''. I sent him the answer (noted at the end of *Some inequalities for ordinals* in this Companion, discussion

concerning p.191), and starting with the 4th printing, he notes this, though in that printing, Solovay's name is misspelled.

[**3a**|*4a*] P.893, end. Add:

14. Let A, B be non-empty sets. Prove that

$$\text{card}(A) \leq \text{card}(B) \quad \text{or} \quad \text{card}(B) \leq \text{card}(A).$$

[Hint: consider the family of pairs (C, f) where C is a subset of A and $f : C \to B$ is an injective map. By Zorn's lemma there is a maximal element. Now finish the proof].

[**4s**] P.895: In title of [Art 24], the first word should end with -ung, not -umg.

[**6a**|*1s*] P.895, reference [Boun 1854]: ''invarialb es'' should be ''invariables''.

[*6a*|**1s**|*4s*] P.899: There are three references [Neu 69]; these should be [Neu 69a], [Neu 69b] and [Neu 69c]. In the second of these, ''*p*-adische'' should be ''*p*-adische**n**''. In reference [Ph 89] part II is shown as ''to appear'', but there is full information for it and for part III on p.360. Reference [Pop 94] should be ''On Grothendieck's conjecture of birational anabelian geometry, *Annals of Math.* (2) **139** (1994) 145-182.''

[**1s**|*4s*] P.901: In [Wit 36] (the first of the two references so numbered in the pre-Springer printings), ''der der'' should be a single ''der'' and ''Charakteristi**c**'' should be ''Charakteristi**k**''. In [Wit 37] (in the pre-Springer printings shown as the second of two references ''[Wi 36]''), ''mit vollkommene**n**'' should be ''mit vollkommene**m**''.

[**4s**] Index: The entry ''algebraically closed'', in addition to p.272, should show pages 178 and 231, while the entry ''algebraic closure'' should refer to only one page, 232. In the entry ''cocycle'', delete the comma after the subheading ''Hilbert's theorem'', since this refers to Hilbert's ''theorem 90'', not to a theorem on p.90.

[**1s**|*4s*] Index: The entry ''adjoint functor'' should be deleted; it does not occur in this book. (The topic is developed in my Math 245 notes.) To the entry for ''discriminant'' add page 270. To the entry for ''*G*-set'' add page 25. In the entry for ''symmetric group'', 29 should be 28, 30.

[**6a**|*1s*] Index: In the entry for ''Jordan-Hölder'', the page number 157 should be 156. In the entry for ''Noetherian'', 407 should be 408.

[**3a**|*4a*] Index: Add to the entry for ''algebraically independent'' the page-number 102. In the entry for ''cyclic extension'' change 388 to 266, 288. Under ''Hilbert'', delete the comma after the subheading ''theorem'', since this again refers to Hilbert's ''theorem 90'', not to a theorem on p.90. The first three page-numbers under ''Riemann surface'' should go instead under ''Riemann-Roch''.

The Axiom of Choice, Zorn's Lemma, and all that

When set theory was formalized in the early 1900's, and a system of axioms set down, it was found (as for Euclidean geometry centuries earlier!) that one of the axioms proposed was not quite as ''obvious'' as the others. This was the axiom that allowed one, in a construction, to simultaneously make an arbitrarily large set of choices. The formal statement was that given a family $(X_i)_{i\in I}$ of nonempty sets, one could choose a family $(x_i)_{i\in I}$ with each $x_i \in X_i$ – in other words, that the product $\Pi_{i\in I} X_i$ of a family of nonempty sets X_i is nonempty!

If the index set I is finite, there is no problem with this: The statement that X_i is nonempty says that there exists (which is all that is meant by ''one can choose'') an element $x_i \in X_i$; a finite number of applications of this observation gives an element $(x_i) \in \Pi_{i\in I} X_i$. (One can make this argument precise using induction.) But some mathematicians argued that if I was infinite, we had no way of knowing that the same was true, so that the general assertion, called the Axiom of Choice, should be rejected. And indeed, some applications of that Axiom led to conclusions that mathematicians found surprising, and possibly unreasonable. (For instance, that there exists a set S of real numbers containing exactly one representative of each coset of the additive subgroup of rational numbers.) We will not go further into the history here. The Axiom is part of the set-theoretic foundations accepted by most mathematicians today, ''Zermelo-Fraenkel set theory with Choice'', abbreviated ZFC.

However, certain reformulations are generally more convenient to use than the Axiom in its raw form. A particularly useful formulation is Zorn's Lemma, stated as part (ii) of the Theorem on the next page, which allows one to carry out constructions that may require infinite sequences of choices, each of which depends on the preceding ones, so that one does not know initially just what choices are to be made and in what order.

In what follows, I assume you are familiar with the concepts of *partially ordered set* and *totally ordered set*. (Many authors shorten one or the other of these terms to *ordered set*; e.g., in Appendix 2.1 Lang [**2**], a partially ordered set is so called.) An element m of a partially ordered set P is called *maximal* if no element of P is strictly greater than m. A subset S of a partially ordered set P is called *bounded* if there exists $t \in P$ such that t is $\geq$ every element of S. A subset S of a partially ordered set P is called a *chain* if it is totally ordered under the induced ordering. An *initial segment* of a chain S means a subset $T \subseteq S$ such that if $u \leq v$ are members of S, and $v \in T$, then $u \in T$.

A *well-ordered* set means a totally ordered set in which every nonempty subset has a *least* element. You should verify the following observations that we will need:

(1) If S is a well-ordered subset of a partially ordered set P, and $t \notin S$ is an upper bound for S in P, then $S \cup \{t\}$ is well-ordered.

(2) If C is a set of well-ordered subsets of a partially ordered set P, such that for all $X, Y \in C$, either X is an initial segment of Y, or Y is an initial segment of X, then $\cup C$ (i.e., $\cup_{X\in C} X$) is well-ordered.

In the Theorem below, we assume the axioms of ZFC *other than* the Axiom of Choice, and sketch a proof that under these assumptions, four statements, one of which is that Axiom, and another of which is Zorn's Lemma, are equivalent. We will not list the other Axioms of ZFC, but simply allow ourselves to use familiar set-theoretic techniques other than infinite systems of choices. For a brief discussion of these axioms see [**1**, §4.4]; for a detailed development, see a text on set theory. In Lang [**2**] the Axiom of Choice is assumed without being stated, so that in his Appendix on Set Theory, statements (ii) and (iii)

below are simply ''proved''. The development here is shorter than Lang's, and explicitly shows the relation with the Axiom of Choice.

To indicate how assertion (ii) below is useful, let me note that in the typical situation where it is to be applied, P represents a set of ''partially completed'' constructions, and the relation $q \geq p$ means ''q is a partial construction which extends (carries further) the partial construction p''. For a given application, one has to set up the P that represents one's method of construction (sometimes tricky), verify that it satisfies the hypotheses of Zorn's Lemma (usually easy), and then show that a *maximal* element of P corresponds to the sort of completed construction one wants (frequently a nontrivial argument in the area of mathematics one is applying the Lemma to). We will see examples of this procedure below in the proofs of the three other conditions from (ii).

Theorem 1. *The following statements are equivalent* (*assuming Zermelo-Fraenkel Set Theory but not the Axiom of Choice*).

(i) (Axiom of Choice) *If* $(X_i)_{i \in I}$ *is a family of nonempty sets, then* $\Pi_I X_i$ *is also nonempty.*

(ii) (Zorn's Lemma) *If* P *is a nonempty partially ordered set with the property that every chain in* P *is bounded, then* P *has a maximal element.*

(iii) (Well-Ordering Principle) *Every set* X *can be well-ordered.* (*I.e., for every* X *there exists a binary relation* $\leq$ *on* X *which makes it a well-ordered set.*)

(iv) (Comparability of cardinalities, or Bernstein's Theorem) *Given any two sets* X *and* Y, *there exists either a bijection between* X *and a subset of* Y, *or a bijection between* Y *and a subset of* X.

Sketch of Proof. We shall first prove (i)$\Rightarrow$(ii) (which is the hard step, and is the important implication for our purposes, since it means that assuming ZFC, we can use Zorn's Lemma), then show that (ii) implies each of (i), (iii) and (iv), next prove (iii)$\Rightarrow$(i), and finally, very roughly, indicate how (iv)$\Rightarrow$(iii) is proved. If each part of this argument were complete, the arguments showing (ii)$\Rightarrow$(i) and (ii)$\Rightarrow$(iii) could be omitted, and we would have a cyclic proof of equivalence of our four conditions. The present arrangement allows us to get the equivalence of (i), (ii) and (iii) without needing the details of (iv)$\Rightarrow$(iii), which we haven't time to go into, and also gives three instructive illustrations of how Zorn's Lemma is used.

(i)$\Rightarrow$(ii). Let P satisfy the hypothesis of (ii). If S is any chain in P, then by assumption, the set of upper bounds to S in P is nonempty; from this it is not hard to see that if S does *not* contain a maximal element of P, the set of upper bounds to S which do not lie in S is nonempty. Let us call the latter set $B(S)$. Assuming (i), let us choose a function φ on the set of all chains in P which do not contain maximal elements of P, associating to each such chain S a member of $B(S)$.

The *idea* of what we want to do next is to start with any element $p \in P$; if it is not maximal, let $p' = \varphi(\{p\}) > p$; if p' is not maximal in P let $p'' = \varphi(\{p, p'\}) > p'$; if p'' is not maximal, let $p''' = \varphi(\{p, p', p''\}) > p''$, and so on. If we do not get a maximal element after finitely many of these steps, we let $p^* = \varphi(\{p, p', \dots, p^{(i)}, \dots\})$; if p^* is not maximal in P, we let $p^{*\prime}$ denote the result of applying φ to the chain gotten by appending p^* to the above chain – ''and so on''! Intuitively, this process can ''go on running'' indefinitely, producing a chain of ever growing cardinality, and would run beyond the cardinality of P if it never terminated; hence it must eventually terminate, which can only happen if we hit a maximal element of P. But we do not have at hand the machinery needed to formalize how to ''set such a construction running'' (though this is developed in [**1**]). What we shall use instead is a definition which *characterizes* the set of chains that *would* have been given by the above construction, if we had been

in a position to formalize it.

Namely, having fixed an element $p \in P$, let C denote the set of subsets $S \subseteq P$ which have the properties:

(a) S is a well-ordered chain in P.

(b) p is the least element of S.

(c) For every proper nonempty initial segment $T \subseteq S$, the least element of $S - T$ is $\varphi(T)$.

We see that C is nonempty, because it contains $\{p\}$. We claim that if S and S' are two members of C, then one of them is an initial segment of the other. Indeed, let R denote the union of all sets that are initial segments both of S and of S'; thus, R is the greatest common initial segment of S and of S'. If R were a *proper* initial segment both of S and of S', then by (c), $\varphi(R)$ would be the least element both of $S - R$ and of $S' - R$; but this would make $R \cup \{\varphi(R)\}$ an initial segment both of S and of S', contradicting the choice of R. So R cannot be proper in both S and in S', i.e., it must equal one of them. That one will thus be an initial segment of the other, as claimed.

Hence if we now denote by U the union of all members of C, then by (2) U is well-ordered. Clearly, also, all members of C are initial segments of U, and the least element of U is p. We claim that U also satisfies (c) above. Indeed, if T is a proper nonempty initial segment of U, there exists some element $u \in U - T$. By construction of U, u belongs to some $S \in C$, and by what we have proved about C, T must be a proper initial segment of S. Hence (c) says that $\varphi(T)$ is the least element of $S - T$, so as S is an initial segment of U, $\varphi(T)$ is also the least element of $U - T$, as required.

This proves that U is itself a member of C. Now suppose U does not contain a maximal element of P. Then $U \cup \{\varphi(U)\}$ will be an element of C (cf. (1) above), which is not a subset of U, contradicting the definition of U. So U must contain a maximal element of P; so such a maximal element exists, establishing (ii).

The above argument was intricate; but the elegance of Zorn's Lemma will be seen in the fact that the proofs of all three other conditions *from* that Lemma are quite direct. I leave it to you to verify that in each case, the partially ordered set described satisfies the hypothesis of Zorn's Lemma, and that a maximal element, whose existence is the conclusion of that Lemma, will be an object of the sort we are trying to get in the case in question. If you have not seen Zorn's Lemma proofs before, I urge you to think through all these verifications carefully!

(ii)$\Rightarrow$(i). Let $(X_i)_{i \in I}$ be given as in (i). We let P be the set of ''partial functions'' on I carrying each $i \in I$ to a member of X_i; formally, we define this to be the set of all subsets $\varphi \subseteq I \times (\cup_I X_i)$ such that for each $i \in I$, any element of φ with first component i has second component in X_i, and for each i there is *at most* one such element. We let P be partially ordered by inclusion (i.e., we consider the partial order relation $\subseteq$ on P). P is nonempty because the *empty* partial function is a member of it; and given any chain S in P, one can verify that the *union* of S (as a family of sets) is an upper bound to S in P. By Zorn's Lemma we conclude that P has a maximal element φ under this ordering; we then verify that such a φ will be a *function*, i.e., a member of $\Pi_I X_i$, as required.

(ii)$\Rightarrow$(iv). (We do this next because the method is very similar to the preceding.) Given sets X and Y, let P be the set of all subsets $\varphi \subseteq X \times Y$ such that for each $x \in X$, φ contains at most one element with first component x, and for each $y \in Y$, φ contains at most one element with second component y, and let us order P by inclusion. As above, Zorn's Lemma gives us a maximal $\varphi \in P$. In this case, we find that maximality implies that *either* φ is a one-to-one function from X to Y *or* the set $\varphi^* = \{(y, x) \mid (x, y) \in \varphi\}$ is a one-to-one function from Y to X; so *one* of these exists, as desired.

(ii)⇒(iii). Given X, we define the set P to consist of all pairs $(A, \preccurlyeq)$ such that A is a subset of X, and $\preccurlyeq$ is a well-ordering of A. We then partially order P by letting $(A, \preccurlyeq) \leq (A', \preccurlyeq')$ if $A \subseteq A'$, $\preccurlyeq$ is the restriction of $\preccurlyeq'$ to A, and A is an *initial segment* of A' under $\preccurlyeq'$. Given a chain $\{(A_i, \preccurlyeq_i) \mid i \in I\}$ in P, one gets an upper bound by taking the union of the A_i, ordered by the union of the $\preccurlyeq_i$. The ordering on this union is a well-ordering by (2). (If in defining the order relation $(A, \preccurlyeq) \leq (A', \preccurlyeq')$ on P, we had left out condition that A be an initial segment of A', this would not be true!) By (ii), P has a maximal element. Now let us note that for any element $(A, \preccurlyeq) \in P$ with A a *proper* subset of X, we can add one more element of X to A to get a larger set A', extend $\preccurlyeq$ to an ordering $\preccurlyeq'$ in which this new element is greater than all others, and in view of (1), the result is an element of P above our given element. Hence for $(A, \preccurlyeq) \in P$ that is *maximal*, we must have $A = X$, so $\preccurlyeq$ is a well-ordering of the whole set X, as required.

(iii)⇒(i). Given $(X_i)_{i \in I}$ as in (i), let us choose, using (iii), some well-ordering $\leq$ of the set $X = \bigcup_I X_i$. For each $i \in I$ let x_i be the *least* element of the nonempty subset $X_i \subseteq X$ under this well-ordering. Then $(x_i)_{i \in I}$ is a member of $\prod_I X_i$.

(iv)⇒(iii) requires a bit of machinery that we will not develop here: Given any set X, one can show that there exists a well-ordered set Y which cannot be put in bijective correspondence with any subset of X. (One constructs this, loosely, by ''welding together'' copies of *all* well-ordered sets whose underlying sets are subsets of X, so that each is embedded in Y as an initial segment, and then adding one more element at the top.) But (iv) tells us that *either* Y can be put in bijective correspondence with a subset of X, or X with a subset of Y. Hence the latter must hold; and the resulting bijection between X and a well-ordered set induces a well-ordering on X, as required. □

I will assume the equivalent statements (i)-(iv) for the remainder of this course. The one we will use by far the most often is Zorn's Lemma.

Let us note that the formulation of Zorn's Lemma is slightly redundant: the empty set $\varnothing$ is a chain in any partially ordered set P, and an upper bound in P to that chain means an arbitrary element of P. Hence if *every* chain, including $\varnothing$, has an upper bound, it is unnecessary to add a separate assumption that P is nonempty. I have stated Zorn's Lemma as I did to conform with common usage, but in fact, you can see that in the proof of (ii)⇒(i) above, the verification of the nonemptiness of P used precisely the element that would have been gotten by applying our construction of upper bounds to the empty chain; and this is also true of the corresponding verifications (not given explicitly) in the proofs of (ii)⇒(iii) and (ii)⇒(iv). In almost every proof using Zorn's Lemma that I have seen, either the same is true, or the proof could be made a bit neater by using that approach! Only very rarely does one need different methods to find upper bounds for nonempty chains and for the empty chain. (However, some authors muddy the waters by defining a ''chain'' to be a *nonempty* totally ordered subset.) In our proof of Zorn's Lemma from the Axiom of Choice, it would likewise have been more elegant not to bring in p, but to replace (b) and (c) by the single condition gotten by removing ''nonempty'' from (c).

References

[**1**] George M. Bergman, *An Invitation to General Algebra and Universal Constructions*, pub. Henry Helson, Berkeley, CA, 1998. ii+398 MR **99h**:18001.

[**2**] Serge Lang, *Algebra*, revised third edition, Springer GTM v.211, 2002.

A principal ideal domain that is not Euclidean

developed as a series of exercises

Let α denote the complex number $(1+\sqrt{-19})/2$, and R the ring $\mathbf{Z}[\alpha]$. We shall show that R is a principal ideal domain, but not a Euclidean ring. This is Exercise III.3.8 of Hungerford's *Algebra* (2nd edition), but no hints are given there; the proof outlined here was sketched for me by H. W. Lenstra, Jr..

(1) Verify that $\alpha^2 - \alpha + 5 = 0$, that $R = \{m+n\alpha \mid m,n \in \mathbf{Z}\} = \{m+n\bar{\alpha} \mid m,n \in \mathbf{Z}\}$, where the bar denotes complex conjugation, and that the map $x \mapsto |x|^2 = x\bar{x}$ is nonnegative integer valued and respects multiplication.

(2) Deduce that $|x|^2 = 1$ for all units of R, and using a lower bound on the absolute value of the imaginary part of any non-real member of R, conclude that the only units of R are ± 1.

(3) Assuming R has a Euclidean function φ, let x be a nonzero nonunit of R minimizing $\varphi(x)$. Show that R/xR consists of the images in this ring of 0 and the units of R, hence has cardinality at most 3. What nonzero rings are there of such cardinalities? Show that $\alpha^2 - \alpha + 5 = 0$ has no solution in any of these rings, and deduce a contradiction, showing that R is not Euclidean.

We shall now show that R *is* a principal ideal domain. To do this, let I be any nonzero ideal of R, and x a nonzero element of I of least absolute value, i.e., minimizing the integer $x\bar{x}$. We shall prove $I = xR$. (Thus, we are using the function $x \mapsto x\bar{x}$ as a substitute for a Euclidean function, even though it does not have all the properties of one.)

For convenience, let us ''normalize'' our problem by taking $J = x^{-1}I$. Thus, J is an R-submodule of $\mathbf{C}$, containing R and having no nonzero element of absolute value < 1. We shall show from these properties that $J - R = \varnothing$, i.e., that $J = R$.

(4) Show that any element of J that has distance less than 1 from some element of R must belong to R. Deduce that in any element of $J-R$, the imaginary part must differ from any integral multiple of $\sqrt{19}/2$ by at least $\sqrt{3}/2$. (Suggestion: draw a picture showing the set of complex numbers which the preceding observation excludes. However, unless you are told the contrary, this picture does not replace a proof; it is merely to help you find a proof.)

(5) Deduce that if $J-R$ is nonempty, it must contain an element y with imaginary part in the range $[\sqrt{3}/2,\ \sqrt{19}/2-\sqrt{3}/2]$, and real part in the range $(-1/2,\ 1/2]$.

(6) Show that for such a y, the element $2y$ will have imaginary part too close to $\sqrt{19}/2$ to lie in $J-R$. Deduce that $y = \alpha/2$ or $-\bar{\alpha}/2$, and hence that $\alpha\bar{\alpha}/2 \in J$.

(7) Compute $\alpha\bar{\alpha}/2$, and obtain a contradiction. Conclude that R is a principal ideal domain.

(8) What goes wrong with these arguments if we replace 19 throughout by 17? By 23?

Lüroth's Theorem, and some related results,
developed as a series of exercises

A *simple transcendental extension* of a field k means an extension of the form $k(t)$ where t is transcendental over k; in other words, an extension isomorphic to the field of rational functions in one indeterminate over k. Throughout this set of exercises, $k(t)$ will denote such a field.

We shall prove that every subextension of $k(t)$, other than k itself, is also a simple transcendental extension of k. We shall need to consider algebraic relations satisfied by elements of $k(t)$ over subfields, hence we shall work with the polynomial ring $k(t)[x]$. In calculating with elements of this ring we shall make use of the subring $k[t, x]$, which is a polynomial ring over k in two indeterminates.

(1) Show that every nonzero element $f \in k(t)[x]$ can be written uniquely $f = P(t,x)/Q(t)$ where $P(t,x) \in k[t, x]$, $Q(t) \in k[t]$, P and Q are relatively prime in the unique factorization domain $k[t, x]$, and Q is monic as a polynomial in t.

From now on, whenever we write an element $f \in k(t)[x]$ as $P(t,x)/Q(t)$, we shall understand P and Q to have the above properties. In this situation, let us define the *height* of f to be the nonnegative integer $\mathrm{ht}(f) = \max(\deg_t(P(t,x)), \deg_t(Q(t)))$, where $\deg_t$ means ''degree as a polynomial in t''.

In particular, if $u \in k(t)$ and we write $u = P(t)/Q(t)$, then the above assumptions on P and Q, and the above definition of the height $\mathrm{ht}(u)$, will be understood to apply.

(2) Suppose $f = P(t,x)/Q(t)$ is *monic* as a polynomial in x over $k(t)$. Show that $\mathrm{ht}(f) = \deg_t(P(t,x))$, and that $P(t,x)$ is not divisible by any nonunit element of $k[t]$.

(3) Deduce that if $f, g \in k(t)[x] - \{0\}$ are both monic as polynomials in x, then $\mathrm{ht}(fg) = \mathrm{ht}(f) + \mathrm{ht}(g)$.

(4) If u is any element of $k(t)$ not in k, show that there is an element u', of the same height as u, such that $k(u') = k(u)$, and with the properties that when u' is written $P'(t)/Q'(t)$ we have $\deg_t(P') > \deg_t(Q')$, and that P' (as well as Q') is monic.

(You will find that u' can be taken to be of the form αu or $\alpha/(u-\beta)$, where $\alpha, \beta \in k$.)

The next observation shows that t, though transcendental over k, is algebraic over any nontrivial intermediate field.

(5) Given any $u = P(t)/Q(t) \in k(t) - k$, verify that t is a root of the polynomial $P(x) - uQ(x) \in k(u)[x]$. Show further that if $\deg_t(P) > \deg_t(Q)$, and P is monic, then the above polynomial is monic.

Now suppose L is any intermediate field: $k \subset L \subseteq k(t)$. Let $u = P(t)/Q(t)$ be an element of $L - k$ chosen to minimize $n = \mathrm{ht}(u)$, and, using (4), also taken so that $\deg_t(P) > \deg_t(Q)$ and P is monic. These conditions will be assumed in the next three steps.

(6) Show that every polynomial $f \in L[x]$ has height either 0 or $\geq n$, and that $P(x) - uQ(x)$ has height exactly n; deduce from (3) that $P(x) - uQ(x)$ is either irreducible in $L[x]$, or divisible by a nonunit element of $k[x]$.

(7) Show that if $P(x) - uQ(x)$ is divisible in $L[x]$ by an element of $k[x]$, then this element must divide both $P(x)$ and $Q(x)$. Deduce that this element must be a unit. (Suggestion for the first assertion: extend $\{1, u\}$ to a basis of L over k. You might prove that assertion in the general context of divisibility of a polynomial over any field L by a polynomial over a subfield k. The second assertion depends on our assumptions on P and Q, specific to our present situation.)

(8) Conclude that $P(x) - uQ(x)$ is the minimal polynomial of t over L. Deduce that $P(x) - uQ(x)$ is also the minimal polynomial of t over $k(u) \subseteq L$, and from this that $L = k(u)$.

(9) Deduce:

Lüroth's Theorem. *If $k(t)$ is a simple transcendental extension of a field k, and $k \subset L \subseteq k(t)$, then L is also a simple transcendental extension of k, and is generated over k by any element $u \in L$ of minimal positive height. Moreover,* $[k(t) : L] = \mathrm{ht}(u)$.

To tie things up, we would like to be able to conclude that for every element $u \in k(t) - k$, $[k(t):k(u)] = \mathrm{ht}(u)$. Unfortunately, we have only proved (8) for elements u having minimal height in an intermediate field L, and nothing we have proved excludes the possibility that if we start with a random element u, $k(u) - k$ might contain an element of smaller height than u. The only way I see to get the desired result is to give a second proof of the irreducibility of $P(x) - uQ(x)$, with the hypothesis of minimal height replaced by the hypothesis $L = k(u)$.

(10) Show for arbitrary $u = P(t)/Q(t) \in k(t) - k$ that $P(x) - uQ(x)$ is irreducible in $k(u)[x]$. (Hint: Apply Gauss's Lemma to $k[u]$.)

(11) Deduce that in the above situation, $[k(t):k(u)] = \mathrm{ht}(u)$, and conclude that u *is* of minimal height in $k(u) - k$.

(12) Deduce the following result, including the ''that is'' clause. (This is essentially Exercise V.2.6(d) of Hungerford; cf. Exercise IV.10 of Lang),

Corollary. *The simple generators of $k(t)$ as an extension field of k are precisely the elements of height* 1, *that is, the elements of the form $(at + b)/(ct + d)$ with $ad - bc \neq 0$. Hence every automorphism of $k(t)$ over k takes t to an element of this form.*

Lüroth's Theorem is analogous to the result that a nontrivial subgroup of the free abelian group $\mathbf{Z}$ of rank 1 is free of rank 1, is generated by any element of minimal absolute value, and has index in $\mathbf{Z}$ equal to this absolute value. However, analogous results do not hold in all situations:

(13) Show that the subalgebra $k[x^2, x^3] \subseteq k[x]$ is not a polynomial ring over k. (Suggestion: show it is not a unique factorization domain.)

If one looks at subfields of fields of rational functions in more than one indeterminate, the analog of Lüroth's Theorem fails – these need not, in general, be isomorphic to rational function fields. However, to give counterexamples, and to develop the partial positive results that do hold, we would need the theory of transcendence degree, and the concept of separability for transcendental extensions (Hungerford, Chapter VI; Lang, Chapter VIII).

Quadratic Reciprocity

developed from the theory of finite fields, as a series of exercises

If p is a prime, then an element $x \in \mathbf{Z}/p\mathbf{Z} - \{0\}$ is called a *quadratic residue* if $(\exists y \in \mathbf{Z}/p\mathbf{Z} - \{0\})\ x = y^2$. One also, loosely, calls an integer q not divisible by p a ''quadratic residue modulo p'' if $\bar{q} \in \mathbf{Z}/p\mathbf{Z}$ is a quadratic residue.

Given p, we can clearly determine with a finite amount of work which elements of $\mathbf{Z}/p\mathbf{Z}$ are quadratic residues, and thus which elements of $\mathbf{Z}$ are quadratic residues modulo p; the latter set will be periodic with period p. But given $q \in \mathbf{Z}$, it is not evident whether there will be any pattern in the set of primes p modulo which q is a quadratic residue. We shall discover below that there is.

We begin the study of this question with a reduction:

(1) Let p be an odd prime, $(\mathbf{Z}/p\mathbf{Z})^*$ the group of units of $\mathbf{Z}/p\mathbf{Z}$, and $(\mathbf{Z}/p\mathbf{Z})^{*2}$ the subgroup $\{x^2 \mid x \in (\mathbf{Z}/p\mathbf{Z})^*\}$. (i) Find $[(\mathbf{Z}/p\mathbf{Z})^* : (\mathbf{Z}/p\mathbf{Z})^{*2}]$. (ii) Show from this that one can determine, given two integers r and s relatively prime to p, whether rs is a quadratic residue modulo p, if one knows whether r is and whether s is. (iii) Deduce that one can describe the set of primes modulo which an arbitrary integer q is a quadratic residue if one can do this whenever q is a prime, and also when $q = -1$.

Until further mention, p and q will be distinct *odd primes*. We shall establish below a relation between the conditions that q be a quadratic residue modulo p, and that p be a quadratic residue modulo q, known as the Law of Quadratic Reciprocity, by studying in *two* ways the question

(*) Does the Galois group of the polynomial $x^q - 1$ over $\mathbf{Z}/p\mathbf{Z}$ act by even permutations of the roots of that polynomial?

and comparing the answers that we get. As noted in (1), we also need to investigate the cases $q = -1$ and $q = 2$. These will be dealt with subsequently.

The first of our approaches to (*) will be based on the fact that the Galois group is generated by the map $a \mapsto a^p$; the second, on the use of the discriminant. For the first approach we begin with a general result on cyclic groups:

(2) If $\langle x \rangle$ is a finite cyclic group, and $y \in \langle x \rangle$, show that the map $z \mapsto zy$ is an even permutation of the elements of $\langle x \rangle$ if and only if y is a square in $\langle x \rangle$.

For p, q still as above, let $(\mathbf{Z}/p\mathbf{Z})(\zeta_q)$ be the extension of $\mathbf{Z}/p\mathbf{Z}$ by a primitive qth root of unity, ζ_q.

(3) Show that the following conditions are equivalent:

(a) The pth-power map acts by an even permutation on $\langle \zeta_q \rangle \subseteq \mathbf{Z}/p\mathbf{Z}(\zeta_q)^*$.

(b) Multiplication by p acts as an even permutation on the elements of $\mathbf{Z}/q\mathbf{Z}$.

(c) p is a quadratic residue modulo q.

We have now related something that happens in characteristic p and something that happens in characteristic q. But (a) is formulated, not in $\mathbf{Z}/p\mathbf{Z}$, but $(\mathbf{Z}/p\mathbf{Z})(\zeta_q)$. The next step is to reduce this to a computation in $\mathbf{Z}/p\mathbf{Z}$.

(4) Compute the discriminant of $x^q - 1$ in $\mathbf{Z}$. (Suggestion: Consider successively the relation among the following elements: $d = \Pi_{i<j} (\zeta_q^i - \zeta_q^j)^2$, $d' = \Pi_{i \neq j} (\zeta_q^i - \zeta_q^j)$, $d'' = \Pi_{i \neq j} (1 - \zeta_q^{j-i})$, $E(1) = \Pi_{i \neq 0} (1 - \zeta_q^i)$, $E(x) = \Pi_{i \neq 0} (x - \zeta_q^i)$. If you do it this way, be more precise than I have been about the ranges of i and j. And be careful!)

(5) Obtain from (4) a necessary and sufficient condition for the Galois group of $(\mathbf{Z}/p\mathbf{Z})(\zeta_q)$ to act by even permutations on the roots of $x^q - 1$. In putting this criterion in the simplest possible form, recall that for nonzero elements a and b of a field, a is a square if and only if ab^2 is a square.

(6) Comparing the results of (3) and (5), get a relation between p being a quadratic residue modulo q and q being a quadratic residue modulo p, which holds whenever $q \equiv 1 \pmod 4$. Reversing the roles of p and q, show that the same relation holds when $p \equiv 1 \pmod 4$.

To handle the remaining case, where both p and q are $\equiv 3 \pmod 4$, we discover that we need to know when -1 is a quadratic residue. This is something we wanted to find out anyway, so

(7) Show that -1 is a quadratic residue modulo p if and only if the group $\mathbf{Z}_p^*$ has an element of order 4. Determine conditions for this to happen in terms of the order of this group, and hence in terms of p.

(8) Complete the incomplete result of (6). Note: this result, the Law of Quadratic Reciprocity, is often stated in terms of the *Legendre Symbol* $(\frac{n}{p})$, which is defined, for p a prime, and n an integer not divisible by p, to be $+1$ if n is a quadratic residue modulo p, -1 if not. The special situation when $p, q \equiv 3 \pmod 4$ is expressed by bringing into the statement of the Law a term $(-1)^{((p-1)/2)((q-1)/2)}$.

As a practical application:

(9) Characterize the set of odd primes p such that -39 is a quadratic residue modulo p as those belonging to an explict list of congruence classes modulo some integer. Give all such primes < 100. (You need *not* find square roots of -39 modulo each!)

It remains to handle the case $q = 2$. Note that all fields $\mathbf{Z}/p\mathbf{Z}$ (p odd) have a primitive square root of unity, namely -1, so the Galois group of $(\mathbf{Z}/p\mathbf{Z})(\zeta_2)$ is trivial; nevertheless, 2 is a quadratic residue in some fields $\mathbf{Z}/p\mathbf{Z}$ but not in others. It turns out that, just as the case $q = -1$ involved the study of primitive 4th roots of 1, so the case $q = 2$ can be solved by considering primitive 8th roots of 1, as is suggested by the formula $\zeta_8 = (1+i)/\sqrt{2}$.

(10) Without assuming the above formula, show that in any field K containing a primitive 8th root of unity ζ_8, the element $\zeta_8 + \zeta_8^{-1}$ is a square root of 2. Taking $K = (\mathbf{Z}/p\mathbf{Z})(\zeta_8)$, deduce that $\mathbf{Z}/p\mathbf{Z}$ contains a square root of 2 if and only if $\zeta_8 + \zeta_8^{-1}$ is fixed under the pth-power automorphism. Determine for what values of p this holds. (Hint: given any field K with a primitive 8th root of unity ζ_8, and an odd integer n, compare $\zeta_8 + \zeta_8^{-1}$ and $\zeta_8^n + \zeta_8^{-n}$.)

This completes the task we set ourselves to begin with. But one may generalize our question, and ask when an element of a larger field $\mathbf{F}_{p^n}$ is a square in that field. This is answered in

(11) Letting $\mathbf{F}_{p^n}$ denote the field of p^n elements, show that the *norm* map: $\mathbf{F}_{p^n} \to \mathbf{Z}/p\mathbf{Z}$ is surjective. (This is essentially Hungerford's Exercise V.7.2.) Deduce that an element $a \in \mathbf{F}_{p^n}$ is a square in this field if and only if its norm in $\mathbf{Z}/p\mathbf{Z}$ is a quadratic residue.

There are other related questions which do not have such neat answers. For instance, given a prime p one may ask whether *neither*, *one* or *both* of $4 \pm \sqrt{3}$ are squares in $(\mathbf{Z}/p\mathbf{Z})(\sqrt{3})$.

(12) Show (with the help of the preceding results) how to answer the above question in terms of congruences satisfied by p, in all cases *except* when 3 and $(4+\sqrt{3})(4-\sqrt{3}) = 13$ are both quadratic residues modulo p.

I am told by a number-theorist that in the case excluded above, the answer is not determined by congruences satisfied by p; however, from the fact that 3 is a quadratic residue mod p, it follows that there is a factorization $p = (a+\sqrt{3}b)(a-\sqrt{3}b)$ in $\mathbf{Z}[\sqrt{3}]$, and the question *can*, it seems, be answered in terms of congruences satisfied by the coefficients a and b in this factorization.

We end with an exercise which is tangential to the above results, and uses only (1) above:

(13) If m and n are integers, show that the polynomial $(x^2-m)(x^2-n)(x^2-mn)$ has a root in every field $\mathbf{Z}/p\mathbf{Z}$. By appropriate choice of m and n, get an example of a polynomial in $\mathbf{Z}[x]$ having a root in every field of finite characteristic, but no root in $\mathbf{Q}$.

Solution in radicals of polynomials of degree ≤ 4

developed as a series of exercises

The idea of these exercises can be motivated as follows. Suppose u and v are the roots of a quadratic equation $x^2 + ax + b = 0$. The quadratic formula expresses these roots as $u = A_0 + A_1$, $v = A_0 - A_1$, where A_0 is a polynomial in the coefficients of the given equation, and A_1 is the square root of such a polynomial; this is clearly consistent with the fact that A_0 is the same in both expressions, while A_1 changes sign. And in fact, if we solve the above pair of equations for A_0 and A_1 and then write out the resulting expressions for A_0 and A_1^2 and compare with the expressions for the coefficients a and b of the given quadratic equation in terms of u and v, we immediately see that we can express A_0 and A_1^2 in terms of a and b, allowing us to rederive the quadratic formula.

Below, we will seek to generalize this idea to get cubic and quartic formulas as well.

Throughout these exercises:

n will denote one of the values 2, 3, 4 (the degree of the equation we seek to solve),

F will be a field which contains primitive mth roots of unity for all $m \leq n$, hence has characteristic either 0 or $> n$,

$u_1, \dots, u_n$ will be elements of F (which we will later take to be roots of a polynomial; but to begin with, they will just be arbitrary elements on which we will perform some interesting manipulations)

ω (when used) will denote a primitive cube root of unity assumed to lie in F,

i (when used) will denote a primitive fourth root of unity assumed to lie in F.

(1) For each of the systems of equations (a)-(d) below, assuming the the existence in F of the roots of unity occurring in those equations, verify that $u_1, \dots, u_n$ can be written in the form shown, for unique $A_0, \dots, A_{n-1} \in F$. You may assume results of undergraduate linear algebra. (Suggestion: Write each of these systems of equations as a matrix equation, and compute the *square* of the matrix in question. You can simplify your computation by noting certain properties of the rows and columns which force most columns to have zero product with most rows. Also recall the condition on the characteristic, mentioned in the definition of F above.)

(a) $n = 2$

$$\begin{aligned} u_1 &= A_0 + A_1 \\ u_2 &= A_0 - A_1. \end{aligned}$$

(b) $n = 3$

$$\begin{aligned} u_1 &= A_0 + A_1 + A_2 \\ u_2 &= A_0 + \omega A_1 + \omega^2 A_2 \\ u_3 &= A_0 + \omega^2 A_1 + \omega A_2. \end{aligned}$$

(c) $n = 4$

$$\begin{aligned} u_1 &= A_0 + A_1 + A_2 + A_3 \\ u_2 &= A_0 + iA_1 - A_2 - iA_3 \\ u_3 &= A_0 - A_1 + A_2 - A_3 \\ u_4 &= A_0 - iA_1 - A_2 + iA_3. \end{aligned}$$

(d) $n = 4$

$$\begin{aligned} u_1 &= A_0 + A_1 + A_2 + A_3 \\ u_2 &= A_0 + A_1 - A_2 - A_3 \\ u_3 &= A_0 - A_1 + A_2 - A_3 \\ u_4 &= A_0 - A_1 - A_2 + A_3. \end{aligned}$$

(2) Suppose we perform a permutation σ on the elements $u_1, \dots, u_n$, writing $u_i' = u_{\sigma(i)}$, and obtain a representation of the new elements u_i' in terms of elements $A_i' \in F$, using the same systems of equations as above. Show that in all four of cases (a-d), we get $A_0' = A_0$, and that in *three* of the four cases, but not in the remaining one, we can say that the elements $A_1', \dots, A_{n-1}'$ are obtained from $A_1, \dots, A_{n-1}$ by a combination of a permutation, and multiplication by certain mth roots of unity, for some $m \leq n$. (Suggestion: *either* verify the result for a certain small number of permutations $\sigma \in S_n$ and show that the whole result follows, *or* note certain conditions characterizing the set of expressions giving the A's, i.e., the rows of your inverse matrix, verify that any permutation of the u's, i.e., the columns, will approximately preserve those conditions, and deduce that such a permutation will have the asserted affect on the A's.)

(3) Deduce that in each of the three ''good'' cases, the $n-1$ elementary symmetric polynomials in A_1^m, ... , A_{n-1}^m can be expressed as polynomials in the n elementary symmetric polynomials in $u_1, \dots, u_n$. Here m, which depends on n, is as in the preceding question.

(4) Deduce from this that for $2 \leq n \leq 4$, we can find formulas for the solution of polynomial equations of degree n if we have such formulas for polynomial equations of degree $n-1$, and can take mth roots (for the same m as in the last two questions). More precisely, you should show that under those assumptions, we can get formulas associating to a polynomial equation of degree n a finite set of elements, possibly more than n of them, among which the n roots must lie.

In the next two parts, you get to put the above theory into practice! Note that though (6) deals with a harder case than (5), it asks for less: only that you go through the reduction, and not that you grind out an explicit formula.

Before doing either or both of these, you might look ahead at (7), and see whether you wish to incorporate the idea thereof into your calculations. Alternatively, you could do (7) afterwards and show retroactively how it can be used to sharpen the results you get in (5) and/or (6).

(5) Carry out the procedure outlined in (1-4) for $n = 3$, culminating in a formula for solving the general monic cubic equation $x^3 + ax^2 + bx + c = 0$.

(6) Carry out the *reduction* outlined above for $n = 4$. That is, derive formulas which can be used to solve a quartic equation $x^4 + ax^3 + bx^2 + cx + d = 0$ in terms of the roots of an appropriate cubic. Suggestion: begin by showing how you can reduce to the case $u_1 + u_2 + u_3 + u_4 = 0$.

We now return to the problem noted in (4), that our procedure gives us a set of ''candidate roots'', which may be larger than the set of actual roots of our polynomial. You might prepare for the next problem by determining how many ''candidate roots'' it gives in each case. Be careful; some cases that may look at first as though they are different will turn out to be the same.

(7) Verify that in the ''good'' cases of (2) arising in the solution of the cubic and quartic equations, one has $A_1' \dots A_{n-1}' = A_1 \dots A_{n-1}$. Translate this into a statement about symmetric polynomials, and show how it allows us to modify our procedures so as to get exactly the set of roots of the given equation.

(8) (Assumes the student has seen a presentation of the *resolvant cubic* of a quartic polynomial; e.g., Hungerford, Lemmas V.4.9 and V.4.10.) In the $n = 4$ case, suppose the coefficients of the given monic quartic equation lie in a field $K \subseteq F$. Show that for A_1, A_2, A_3 defined as in (1)(d), the extension fields $K(A_1^2)$, $K(A_2^2)$, $K(A_3^2)$ are the fields generated over K by the three roots of the resolvant cubic of the given quartic.

Infinite Galois theory, Stone spaces, and profinite groups

0. Prerequisites, and notation. In addition to elementary Galois theory, this note assumes familiarity with the concept of inverse limit, and the particular example of the ring of p-adic integers. Starting in §2, basic point-set topology is also assumed.

We shall regard elements of inverse limit sets $\varprojlim_{i \in I} X_i$ as families $(x_i)_{i \in I}$ $(x_i \in X_i)$ satisfying the conditions

$$f_{i,i'}(x_i) \;=\; x_{i'},$$

where the $f_{i,i'}$ are the maps of the given inverse system. We will call the above equations the ''compatibility conditions''.

1. Infinite Galois groups. The Galois group of an infinite-dimensional normal algebraic extension is easily described in terms of the Galois groups of its subextensions of finite degree.

Theorem 1. *Suppose E/K is a normal algebraic field extension, possibly of infinite degree. Let I be the set of intermediate fields normal of finite degree over K. If $F_2 \subseteq F_1$ are members of I, let f_{F_1,F_2}: $\mathrm{Aut}_K F_1 \to \mathrm{Aut}_K F_2$ denote restriction of automorphisms from F_1 to F_2. Then $\mathrm{Aut}_K E$ (regarded as mapped into each of these groups by restriction) may be identified with the inverse limit of the system of finite groups $\mathrm{Aut}_K F$ $(F \in I)$ and surjective homomorphisms f_{F_1,F_2}.*

More generally, the same conclusion holds with I replaced by any subset $J \subseteq I$ which has union E, and is directed under inclusion (equivalently, inversely directed under the ordering of I by reverse inclusion. Which ordering is preferable depends on the precise formalism one adopts for inverse limits).

Sketch of Proof. Let J be as in the second paragraph.

Given $\sigma \in \mathrm{Aut}_K E$, restriction of σ to the subfields $F \in J$ yields a system of elements of the groups $\mathrm{Aut}_K F$ satisfying the compatibility conditions for an element of the inverse limit group; thus we get a natural homomorphism $\mathrm{Aut}_K E \to \varprojlim_{F \in J} \mathrm{Aut}_K F$. Because $E = \cup_{F \in J} F$, this homomorphism is one-to-one.

To see that it is onto, let us be given an element $(\sigma_F)_{F \in J}$ $(\sigma_F \in \mathrm{Aut}_K F)$ of the inverse limit. For each $x \in E$, let $\sigma(x)$ denote the common value of $\sigma_F(x)$ as F ranges over all $F \in J$ which contain x. These values are indeed all equal, because given $F_1, F_2 \in J$ both containing x, there will exist $F_3 \in J$ containing F_1 and F_2, and the compatibility conditions on $(\sigma_F)_{F \in J}$ give $\sigma_{F_1}(x) = \sigma_{F_3}(x) = \sigma_{F_2}(x)$. One verifies that σ is a homomorphism using the fact that for any $x, y \in E$ we can find an $F \in J$ with $x, y \in F$; the conditions $\sigma(x+y) = \sigma(x) + \sigma(y)$ etc. then follow from the same conditions on σ_F. Thus, σ is an element of $\mathrm{Aut}_K F$, and it clearly maps to the given element $(\sigma_F)_{F \in J}$ of the inverse limit. □

Example 1. Let $K = \mathbf{Q}$ and let E be the extension gotten by adjoining to $\mathbf{Q}$ the square roots of all positive integers, equivalently, of all primes. Let us write this as the union of the chain of extensions

$$\mathbf{Q} \;\subseteq\; \mathbf{Q}(\sqrt{2}) \;\subseteq\; \mathbf{Q}(\sqrt{2}, \sqrt{3}) \;\subseteq\; \mathbf{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}) \;\subseteq\; \ldots .$$

Now consider the extension of $\mathbf{Q}$ gotten by adjoining the square roots of finitely many primes $p_1, \ldots, p_n$. An automorphism of this field can take each of those square roots to itself or its negative; thus we get an embedding of the Galois group of each such finite extension in $(\mathbf{Z}/2\mathbf{Z})^n$. This embedding is, in fact, an isomorphism. (Sketch of proof: any proper subgroup of $(\mathbf{Z}/2\mathbf{Z})^n$ is annihilated by some $(\mathbf{Z}/2\mathbf{Z})$-linear

functional. If the image of our Galois group in $(\mathbf{Z}/2\mathbf{Z})^n$ were annihilated by a functional having coordinates $(\varepsilon_1, \dots, \varepsilon_n)$ where each $\varepsilon_i = 0$ or 1, then the fixed field of the Galois group would contain $(p_1^{\varepsilon_1} \dots p_n^{\varepsilon_n})^{1/2}$, but this is not a rational number. For a much more general result see [**4**, §VI.8].) Hence $\mathrm{Aut}_{\mathbf{Q}} E$ is the inverse limit of a system of group homomorphisms

$$\dots \to (\mathbf{Z}/2\mathbf{Z}) \times (\mathbf{Z}/2\mathbf{Z}) \times (\mathbf{Z}/2\mathbf{Z}) \to (\mathbf{Z}/2\mathbf{Z}) \times (\mathbf{Z}/2\mathbf{Z}) \to \mathbf{Z}/2\mathbf{Z} \to \{0\},$$

where each map chops off the last coordinate. This inverse limit is an infinite product of $(\mathbf{Z}/2\mathbf{Z})$'s.

Example 2. Again let $K = \mathbf{Q}$. Let p be any prime, and now let E be the union of the chain

$$\mathbf{Q} \subseteq \mathbf{Q}(\zeta_p) \subseteq \mathbf{Q}(\zeta_{p^2}) \subseteq \mathbf{Q}(\zeta_{p^3}) \subseteq \dots,$$

where ζ_n denotes a primitive nth root of unity. Now $\mathrm{Aut}_{\mathbf{Q}} \mathbf{Q}(\zeta_{p^i})$ is isomorphic to the group of units of the ring $\mathbf{Z}/p^i\mathbf{Z}$ [**2**, Proposition V.8.3(iii)] [**4**, Theorem VI.3.1], and we see that the restriction maps among these automorphism groups are induced by the ring homomorphisms $\mathbf{Z}/p^i\mathbf{Z} \to \mathbf{Z}/p^{i-1}\mathbf{Z}$. We deduce that $\mathrm{Aut}_{\mathbf{Q}} E$ is isomorphic to the group of units of the inverse limit of this system of ring homomorphisms, i.e., the units of the ring of p-adic integers.

Example 2′. Take E as above, but take $K = \mathbf{Q}(\zeta_p)$. Then we find that $\mathrm{Aut}_K E$ corresponds to the multiplicative group of p-adic integers congruent to $1 \bmod p$; a subgroup of the full group of units.

Example 3. Let p be a prime, and let k be a field containing primitive p^ith roots of unity for all i. Let K be a pure transcendental extension $k(x)$, and let E be the union of the chain

$$k(x) \subseteq k(x^{1/p}) \subseteq k(x^{1/p^2}) \subseteq k(x^{1/p^3}) \subseteq \dots.$$

The Galois group of the ith term of this chain, $\mathrm{Aut}_{k(x)} k(x^{1/p^i})$, is isomorphic to the *additive* group $\mathbf{Z}/p^i\mathbf{Z}$ [**2**, Theorem V.7.11] [**4**, Theorem VI.6.2(ii)], and the restriction maps are again the obvious ones, so $\mathrm{Aut}_K E$ is isomorphic to the additive group of the p-adic integers.

We note briefly an example of a noncommutative infinite Galois group, leaving the details to the interested reader.

Example 4. Let p be a prime, let $K = \mathbf{Q}(x)$ with x an indeterminate, and let E be obtained by adjoining, for each i, both a primitive p^ith root of unity and a p^ith root of x. It is not hard to show that $\mathrm{Aut}_K E$ is generated by a copy of the multiplicative group of units of the p-adic integers (permuting the roots of unity) and a copy of the additive group of all p-adic integers (acting on the fractional powers of x), and has the structure of the *semidirect product* of these groups, determined by the natural action of the former group on the latter by multiplication in the ring of p-adic integers; equivalently, it is isomorphic to the group of matrices of the form $\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right)$ where a is a unit and b an arbitrary element of that ring.

Our last example, again abelian, is

Example 5. Let p be a prime, let K be the field of p elements, and let E be the algebraic closure of K. Then the subfields of E of finite degree over K are the fields of p^n elements, as n ranges over the positive integers. Each of these is Galois over K, with Galois group generated by the map $a \mapsto a^p$, which on the field of cardinality p^n has order n, so the Galois groups are the groups $\mathbf{Z}/n\mathbf{Z}$ [**2**, Theorem V.5.3] [**4**, Theorem V.5.4]. The restriction homomorphisms are the natural maps $\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z}$ $(m \mid n)$. $\mathrm{Aut}_K E$ is thus the inverse limit of these groups and maps, i.e., the additive

group of the inverse limit of the rings $\mathbf{Z}/n\mathbf{Z}$. That inverse limit can be shown to be the direct product, over all primes p, of the rings of p-adic integers. (Idea: each $\mathbf{Z}/n\mathbf{Z}$ is a direct product of rings $\mathbf{Z}/p^i\mathbf{Z}$.) Thus, our Galois group can be described as the direct product over all primes p of the additive group of p-adic integers.

2. Intermediate fields and closed subgroups. For E/K an infinite algebraic extension we can, as in the finite case, associate to each intermediate field L the subgroup $L^* \subseteq \mathrm{Aut}_K E$ of automorphisms that fix all elements of L, and to each subgroup $H \subseteq \mathrm{Aut}_K E$ its fixed field $H^* \subseteq E$. If E is Galois (i.e., separable and normal) over K, then every intermediate field L is still uniquely determined by the group L^*; that is, $L = L^{**}$ [**2**, Theorem V.3.12] [**4**, Theorem VI.1.2].

On the other hand, a *subgroup* $H \subseteq \mathrm{Aut}_K E$ is in general no longer determined by its fixed field H^*; i.e., H^{**}, which has the same fixed field, can be strictly larger than H. For instance, in the situation of Example 1, let H be the group consisting of all those automorphisms that change the signs of square roots of only *finitely* many primes. This is a proper subgroup, but its fixed field is the same as that of the whole group, since any element $u \in E$ involves square roots of only finitely many primes, hence if u is moved by a member of the whole automorphism group, i.e., is not in $\mathbf{Q}$, it will be moved by some member of H. Thus, here $H^{**} = K^* = \mathrm{Aut}_K E$.

In the situation of Example 3, to get a similar phenomenon, let α be the automorphism of E taking x^{1/p^n} to $\zeta_{p^n} x^{1/p^n}$ for all n (where we assume that our roots of unity and of x are labeled so that $(\zeta_{p^n})^p = \zeta_{p^{n-1}}$ and $(x^{1/p^n})^p = x^{1/p^{n-1}}$), and let $H = <\alpha> \subseteq \mathrm{Aut}_K E$. We find that the restriction of α to each field $K(x^{1/p^n})$ is a generator of the whole Galois group of that field, hence α has no fixed elements in any of these subfields other than elements of K, which means it has no fixed elements in E other than elements of K. So $<\alpha>^{**} = K^* = \mathrm{Aut}_K E$, though the inclusion of $<\alpha>$ in $\mathrm{Aut}_K E$ is certainly proper: it is essentially the inclusion of $\mathbf{Z}$ in the p-adic integers. One can get infinite cyclic subgroups with similar properties in Example 2′ when $p \neq 2$, and in Example 5; in the first case using the automorphism carrying ζ_{p^n} to $(\zeta_{p^n})^{1+p}$ for all n, and in the second, the pth power map (the Frobenius automorphism). (In each of these cases one can, of course, also get cases of $H^{**} \neq H$ in which H^{**} is a proper subgroup of $\mathrm{Aut}_K E$. You might try to verify that for each prime p_0, there is a cyclic subgroup H of the group of Example 5 such that H^{**} is precisely the additive group of p_0-adic integers, regarded as a subgroup of the direct product group of that example.)

Clearly, what is happening in the above examples is that the subgroup H has the same image in each of the finite groups $\mathrm{Aut}_K F$ as does the whole group $\mathrm{Aut}_K E$. In other words, every element of $\mathrm{Aut}_K E$ can be *approximated*, in terms of its behavior on each *finite* subextension, by elements of H. This formulation in terms of approximation suggests that a topology is involved, and indeed it is:

Theorem 2. *Let E/K be a normal algebraic field extension, possibly of infinite degree, let J be any system of normal intermediate fields of finite degree over K which is directed under inclusion and has union E, and let H be any subgroup of $\mathrm{Aut}_K E$. Then the following subgroups of $\mathrm{Aut}_K E$ are equal.*

(i) *H^{**}, i.e. the group of all elements of $\mathrm{Aut}_K E$ fixing the fixed field H^* of H.*

(ii) *The group of all elements of $\mathrm{Aut}_K E$ whose images in each of the finite groups $\mathrm{Aut}_K F$ (under the restriction maps, which we shall write $\pi_F\colon \mathrm{Aut}_K E \to \mathrm{Aut}_K F$) are also images of elements of H; we may write this subgroup loosely as*

$$\varprojlim_{F \in J} \pi_F(H) \;\subseteq\; \varprojlim_{F \in J} \mathrm{Aut}_K F \;=\; \mathrm{Aut}_K E.$$

(iii) *The closure of* H *in* $\mathrm{Aut}_K E$ *under the topology on this group having for a basis of open sets the inverse images, under the restriction maps* π_F, *of the singleton subsets of the groups* $\mathrm{Aut}_K F$ $(F \in J)$*; equivalently, the topology having for basis of open sets the sets* $\{\alpha \in \mathrm{Aut}_K E \mid \alpha(a) = b\}$ $(a, b \in E)$.

Sketch of Proof. (ii) $\subseteq$ (i) is immediate (cf. the examples discussed above). Conversely, consider an element $\alpha \in \mathrm{Aut}_K E$ whose restriction to some intermediate field F of finite degree is *not* equal to the restriction of any element of H. The Fundamental Theorem of Galois Theory, applied to F/K, implies that there is some $a \in F$ fixed by the restrictions of elements of H but not by the restriction of α. This will be an element of H^*, hence as it is not fixed by α, we have $\alpha \notin H^{**}$.

The equivalence of (ii) with (iii), using the first description of the topology on $\mathrm{Aut}_K E$ in (iii), is immediate. To see the equivalence of this with the other description of the topology, note on the one hand that for any $a, b \in E$, if we take $F \in J$ such that $a, b \in F$, then $\{\alpha \in \mathrm{Aut}_K E \mid \alpha(a) = b\}$ will be the inverse image under π_F of a subset of $\mathrm{Aut}_K F$. In the other direction, given $F \in J$, the Theorem of the Primitive Element [**2**, Proposition V.6.15] [**4**, Theorem V.4.6] allows us to write $F = K(a)$, and the inverse image under π_F of any $\alpha_0 \in \mathrm{Aut}_K F$ can now be described as $\{\alpha \in \mathrm{Aut}_K E \mid \alpha(a) = \alpha_0(a)\}$. □

Thus, we get a bijective correspondence between intermediate fields, and subgroups of $\mathrm{Aut}_K E$ closed in the above topology.

The next Lemma, though we will apply it to Galois groups, is stated for general inverse limits of finite sets, so we need to make an explicit choice of how we will index objects of an inversely directed system (which we avoided doing in the last sentence of Theorem 1). From this point on, therefore, let us understand that in such a system, the maps $f_{j,j'}: X_j \to X_{j'}$ are defined for $j \leq j'$ (as in [**1**] and [**5**], but in contrast to the convention of [**4**]).

Lemma 3. *If* X *is the inverse limit of an inversely directed system of finite sets* $(X_j)_{j \in J}$ *and surjective set-maps* $X_{j,j'}: X_j \to X_{j'}$ $(j \leq j')$, *then the canonical maps* $\pi_j: X \to X_j$ *are also surjective.*

Sketch of Proof. Naively, we might hope to show this by starting with any $x_0 \in X_{j_0}$ $(j_0 \in J)$, applying Zorn's Lemma to partial families $(x_j)_{j \in J'}$ with $j_0 \in J' \subseteq J$ that satisfy the compatibility conditions $f_{j,j'}(x_j) = x_{j'}$ $(j \leq j' \in J')$ to get a maximal such family, and hoping that this will be an element of X mapping to x_0.

Unfortunately, the above compatibility conditions on partial families are not strong enough to guarantee that these can be extended to compatible families whose domain is all of J; hence the maximal families that Zorn's Lemma gives us may not have $J' = J$. To get around this, let us call a family $(x_j)_{j \in J'}$ ''robustly compatible'' if it has the property that for any finite set of elements $j_1, \ldots, j_n \in J'$, there exists a $j \in J$ less than or equal to all of these elements, and a $y \in X_j$ such that $f_{j,j_i}(y) = x_{j_i}$ for $i = 1, \ldots, n$. This is easily seen to imply the ordinary compatibility condition noted in the preceding paragraph.

We now claim that any robustly compatible family $(x_j)_{j \in J'}$ can be extended so as to adjoin to J' any $j_0 \in J - J'$. Indeed, given $(x_j)_{j \in J'}$ and $j_0 \notin J'$, suppose that for each $x \in X_{j_0}$, the assignment $x_{j_0} = x$ failed to give a robustly compatible family. Then for each such x we could choose a finite set of elements of J' for which the existence statement required by robust compatibility was false. Let us now apply to the union of these finitely many finite sets the robust compatibility assumption on $(x_j)_{j \in J'}$. This gives a certain $j \in J$ and $y \in X_j$. By the inverse directedness of J we can choose a $j' \in J$ that is $\leq j_0$ and also $\leq j$; by surjectivity of $f_{j',j}$ we can lift y to an element $y' \in X_{j'}$; and if we now let $x = f_{j',j_0}(y') \in X_{j_0}$, we get a contradiction to our choice of the finite subset of J' that allegedly gave a

counterexample to the robust compatibility of the assignment $x_{j_0} = x$.

Hence the Zorn's Lemma argument sketched in the first paragraph, which fails for families that are merely compatible, works for robustly compatible families, completing our proof. □

Recall that every *finite* group can be represented as the Galois group of some Galois field extension of finite degree. There is a corresponding result for infinite Galois extensions.

Lemma 4. *Let G be a topological group. Then the following conditions are equivalent.*

(i) *G is isomorphic as a topological group to the inverse limit of an inversely directed system of finite groups and surjective group homomorphisms, topologized by taking for a basis of open sets the inverse images of singleton subsets of these groups.*

(ii) *G is isomorphic as a topological group to the Galois group of a Galois* (*i.e., separable normal algebraic*) *field extension E/K, topologized by taking for a basis of open sets the sets $\{\alpha \in G \mid \alpha(a) = b\}$ $(a, b \in E)$.*

Proof. We have seen that (ii)⇒(i). Suppose conversely that G is the inverse limit of a system of groups $(G_j)_J$ as in (i). Let us represent each G_j as a group of permutations of a finite set X_j, taking these sets to be disjoint. Since G has a natural map to each G_j, it acquires a natural action on each X_j, hence it acts on their union, $X = \cup X_j$. Let k be any field disjoint from the set X, make X a set of indeterminates over k, let E be the field of rational functions $k(X)$, let G act on E by its permutation action on the indeterminates, and let K be the fixed field of this action. Finally, for each $j \in J$, let $F_j = K(X_j) \subseteq E$.

Since the X_j are disjoint, it is not a priori evident that there are *any* inclusion relations among the fields F_j. We claim, nevertheless, that if $j \leq j'$ in J, then $F_j \supseteq F_{j'}$. To see this, consider the action of G on the compositum $F_j F_{j'} \subseteq E$. This compositum is generated over K by $X_j \cup X_{j'}$. The kernel of the action of G on this union is also the kernel of its action on the subset X_j, since the map of G into $G_{j'}$ factors through the map into G_j; so the groups of automorphisms of F_j and of $F_j F_{j'}$ induced by the actions of G can both be identified with G_j. Further, the fixed field of the action of G on each of these fields is K, since K is the fixed field of its action on E, and by construction both subfields contain K. Hence $[F_j : K] = |G_j| = [F_j F_{j'} : K]$, so the latter field cannot be strictly larger than the former, proving $F_j \supseteq F_{j'}$. Assertion (ii) now follows by the last paragraph of Theorem 2, applied to the system of subfields F_j. □

(In the above proof, we used Lemma 3 to conclude that G mapped surjectively to G_j, so that we could identify its action on X_j with that of G_j. However, if we had not proved that Lemma, we could actually have given the same argument, referring to $\pi_j(G)$ instead of G_j.)

Most of the further basic results of finite Galois theory go over to the infinite case. The main adjustment needed is that when one equates indices of subgroups with degrees of extensions, one does *not* have equalities of cardinality when these are infinite, but merely the statement that one is infinite if and only if the other is. I leave the formulation and verification of these results to the interested reader, noting only a key step that is useful in reducing results on index and degree to the finite case:

Lemma 5. *Let G be an inverse limit of finite groups $(G_j)_{j \in J}$, with projection maps written $\pi_j : G \to G_j$. Let $H_1 \subseteq H_2$ be subgroups of G, with H_1 closed in the topology described above, and let n be a positive integer. Then $[H_2 : H_1] \geq n$ if and only if there exists some $j \in J$ with $[\pi_j(H_2) : \pi_j(H_1)] \geq n$. Moreover, if this holds, then $[\pi_{j'}(H_2) : \pi_{j'}(H_1)] \geq n$ for all $j' \leq j$.*

Idea of Proof. From the definition of our topology it follows that any element not in H_1 has image outside the image of H_1 in some G_j. It follows that two elements in different right cosets of H_1 have images in different right cosets of the image of H_1 in some G_j. This property of G_j is preserved on going to smaller j, and using the inverse directedness of J, one can deduce that elements in n different right cosets of H_1 have images in n different right cosets of the image of H_1 in some G_j. □

3. Stone spaces. What topological groups do in fact occur as inverse limits of systems of finite groups?

Recall that a *topological group* means a group given with a topology with respect to which the group operations are continuous. Most authors also require that the topology be Hausdorff, but we shall not make this a part of our definition, because a stronger condition will soon be imposed. Actually, we have not yet proved that the group operations of our inverse limit groups *are* continuous in our topology, but we shall see this soon. Let us first note some basics.

Lemma 6. *Any inversely directed system of topological spaces $(X_i)_I$ has an inverse limit X, which has for underlying set the inverse limit of the underlying sets of the spaces X_i, and for a basis of open sets the inverse images of all open sets of the spaces X_i (or, alternatively, the inverse images of the members of any family of bases of open sets of the X_i).*

Any inversely directed system of topological groups $(G_i)_I$ has an inverse limit G. Its underlying set is the inverse limit of the underlying sets of the G_i, and its topology and group structure respectively are those of these inverse limits of these objects in the categories of topological spaces and of groups.

Sketch of Proof. It is straightforward to verify that the set-theoretic inverse limit of the X_i together with the indicated topology has the universal property of an inverse limit of the given topological spaces. Let us note why in our description of the inverse limit group G, the group multiplication $G \times G \to G$ is continuous in the inverse limit topology. By the universal property of that topology, it suffices to show that for every $i \in I$, the map $G \times G \to G_i$ gotten by following the group multiplication of G by the projection $\pi_i\colon G \to G_i$ is continuous. But this map can be gotten by projecting the two G's on the left to G_i and applying the multiplication operation of G_i, which is continuous by hypothesis. Continuity of the inverse operation is established in the same way.

It is now straightforward to verify that G, so topologized, has the universal property of an inverse limit of the given groups. □

We now see that the topology introduced in the preceding section is that of the inverse limit of the *discrete* topologies on the given finite groups; the discrete topology being, indeed, the only Hausdorff topology a finite set admits.

Let us note a few more facts that we will need about inverse limits of topological spaces. Recall that a topological space is called *totally disconnected* if for distinct points $x, y \in X$, there exists an *open-closed* subset $U \subseteq X$ containing x but not y.

Lemma 7. *Let X be the inverse limit of an inversely directed system (X_i) of topological spaces.*

The topology on X is the topology induced by regarding X as a subset of $\prod X_i$ with the product topology.

If all X_i are Hausdorff, then X is Hausdorff, and forms a closed subspace of $\prod X_i$.

If all X_i are compact and Hausdorff, then X is compact and Hausdorff.

If all X_i are totally disconnected, then X is totally disconnected.

Sketch of Proof. The first assertion is deduced by noting that the cylinders which form a basis of the topology of $\Pi\, X_i$ intersect X in sets which we have seen form a basis of the inverse limit topology.

That Hausdorffness carries over to products, and hence to subspaces of products, is clear. To show that in the Hausdorff case X is closed in $\Pi\, X_i$, let $x \in (\Pi\, X_i) - X$. Then for some $i < j$ we have $f_{i,j}\pi_i(x) \neq \pi_j(x)$. If we take disjoint open sets U and V in X_j containing $f_{i,j}\pi_i(x)$ and $\pi_j(x)$ respectively, then the intersection of the inverse images in $\Pi\, X_i$ of $f_{i,j}^{-1}(U) \subseteq X_i$ and $V \subseteq X_j$ will be an open subset containing x but disjoint from X, as required.

To see that compactness carries over from the X_i to X when the former are Hausdorff, we may use Tychonoff's Theorem [**3**, Theorem 5.13], which says that compactness carries over to $\Pi\, X_i$, together with the fact that a closed subspace of a compact space is compact.

The assertion on total disconnectedness is immediate. □

Let us now prove

Theorem 8. *A topological space can be written as an inverse limit of finite discrete spaces if and only if it is totally disconnected, compact, and Hausdorff.*

Proof. ''$\Rightarrow$'' follows from the preceding Lemma, since a finite discrete space is totally disconnected compact and Hausdorff.

To prove the reverse implication, let X be a totally disconnected compact Hausdorff space. Let I be the set of *partitions* of X into disjoint open-closed subsets, where we understand such a partition to mean a set of nonempty pairwise disjoint open-closed subsets of X having union X. Let us partially order I by writing $i < j$ if the partition i is a *refinement* of the partition j, i.e., if every member of i is a subset of a member of j. Given arbitrary $i, j \in I$, note that the set of nonempty intersections of members of i with members of j gives a partition of X into open-closed sets which refines both i and j; so the partially ordered set I is inversely directed.

Every member of I is finite because X is compact; let us regard each such i as a finite discrete topological space. If $i < j$, define $f_{i,j}\colon i \to j$ to take each member of i to the member of j which contains it. Let us likewise map X into each $i \in I$ by sending each $x \in X$ to the member of i in which it lies. We note that these maps $X \to i$ are continuous because the members of i are open in X. Since these maps are compatible with the $f_{i,j}$, they induce a continuous map $X \to \varprojlim_I i$.

To see that this map is surjective, let us write any $y \in \varprojlim_I i$ as a family $(U_i)_{i \in I}$, noting that for each $i \in I$, $U_i \in i$ is a nonempty open-closed subset of X. Using the compatibility condition on this member of the inverse limit, we can see that any finite subset of the U_i's has nonempty intersection in X; hence by compactness of X, they all have nonempty intersection. A point of this intersection will be an element of X mapping to y. To see that our map is one-to-one, note that the total disconnectedness condition on X says that any two points of X are distinguished by the map into some two-set partition.

Now our map $X \to \varprojlim_I i$, being a continuous bijection from a compact Hausdorff space to a compact Hausdorff space, is a homeomorphism [**3**, Theorem 5.8], completing the proof of ''$\Leftarrow$''. □

Totally disconnected compact Hausdorff spaces are often called *Stone spaces* after Marshall Stone (who himself called them *Boolean spaces*, for good reasons which we can't go into here; cf. [**3**, Exercise 5.S]). Lemma 6 and Theorem 8 imply in particular that every inverse limit of *finite groups* has a Stone space for underlying topological space.

4. Profinite groups. Is the converse true – given a topological group G whose underlying space is a Stone space, can we write G as an inverse limit of finite groups? To do this, we would have to find, among the finite discrete spaces that are continuous images of G, enough that admit group operations compatible with that of G to express G as their inverse limit. Since these discrete spaces correspond to partitions of G into open-closed sets, we need to study the relation between partitions and operations on sets.

Definition 9. *Given a map of sets,*

$$\alpha\colon\ X^{(1)}\times\ldots\times X^{(n)}\to\ Y, \tag{10}$$

and given a partition $p^{(m)}$ of one of these sets $X^{(m)}$ and a partition q of Y, we shall say that $p^{(m)}$ is compatible *with the mth argument of the map* (10) *and with the partition q if, for any two n-tuples $(x^{(1)},\ldots,x^{(m)},\ldots,x^{(n)})$, $(x^{(1)},\ldots,x'^{(m)},\ldots,x^{(n)})\in X^{(1)}\times\ldots\times X^{(n)}$, differing only in that the mth coordinate of the first is replaced in the second by an element in the same member of $p^{(m)}$, the images of these n-tuples under α lie in the same member of q.*

(*This use of the word ''compatibility'' should not be confused with our use of that term for the conditions relating the components of an inverse limit.*)

If in the above situation we are given a partition $p^{(m)}$ of $X^{(m)}$ for each m, it is easy to see that a necessary and sufficient condition for there to exist a map β making a commuting diagram

$$\begin{array}{ccc} X^{(1)}\times\ldots\times X^{(n)} & \xrightarrow{\ \alpha\ } & Y \\ \downarrow & & \downarrow \\ p^{(1)}\times\ldots\times p^{(n)} & \xrightarrow{\ \beta\ } & q \end{array}$$

is that for each m, $p^{(m)}$ is compatible with the mth argument of α and the partition q of Y.

Let us also note that given any map α as in (10) and any integer $m\le n$, there is a unique *coarsest* partition $p^{(m)}$ on $X^{(m)}$ that is compatible with the mth argument of (10) and the partition q, namely the partition that puts two elements $x^{(m)}$ and $x'^{(m)}$ in the same class if and only if every two elements of $X^{(1)}\times\ldots\times X^{(n)}$ that differ only by replacing $x^{(m)}$ by $x'^{(m)}$ in the mth coordinate have images under (10) in the same member of q.

Now consider a map (10) in which the $X^{(m)}$ and Y are all the same set X, and in which we take the same partition p for all the $p^{(m)}$ and also for q. If p is compatible with each argument of (10) and with the partition p, let us say it is *self-compatible* under (10). The next Lemma reformulates our problem in these terms.

Lemma 11. *Let X be a Stone space* (*a totally disconnected Hausdorff space*), *and $\alpha\colon X^n\to X$ a continuous n-ary operation on X. Then the following conditions are equivalent.*

(i) *X is the inverse limit of an inversely directed system of finite discrete sets X_i with n-ary operations.*

(ii) *For every two distinct points $x, x'\in X$, there exists a partition p of X into open-closed subsets which is self-compatible under α, and such that x and y lie in distinct members of p.*

(iii) *Every partition of X into open-closed subsets can be refined to a partition into open-closed sets which is self-compatible under α.*

More generally, given a finite family of operations on X, the same equivalences hold, with self-

compatibility under all these operations replacing self-compatibility under α.

Sketch of Proof. Clearly (iii)⇒(ii). (ii)⇒(i) is proved in essentially the same way we proved Lemma 8, by noting that the partitions in question form an inversely directed system, and showing that the natural homomorphism from X to the inverse limit of that system is a homeomorphism.

To prove (i)⇒(iii), suppose X is the inverse limit of a system of finite sets X_i with n-ary operations, and p any partition of X into open-closed sets. Each member of p, being open, is a union of members of our basis of the inverse limit topology on X, that is, inverse images of points of various of the X_i. Since X is compact, we can find a *finite* family of these basic open sets, each wholly contained in some member of p, which cover X. To each open set U in this finite covering, let us associate an index i such that U is the inverse image of a point of X_i. If we take an index j which is $\leq$ each of these finitely many indices i, we find that each member of p is a union of inverse images of points of X_j, and since X_j is a homomorphic image of X as a set with n-ary operation α, the partition of X into these inverse images is self-compatible under α.

The case where there are more than one operations works the same way; we gave the arguments in the one-operation case merely for verbal convenience. □

Thus, for a given Stone space with algebraic structure, the question of whether (i) holds translates into that of whether (iii) holds. As a first step, we may ask to what extent, in the general context of (10), the passage from a partition q on Y to the coarsest partition compatible with q and with one of the arguments of α takes ''good'' partitions to ''good'' partitions. This, of course, depends on the sense of ''good''; for a negative example, it does not in general take finite partitions to finite partitions. This can be seen from the set-map

$$\alpha\colon\ \mathbf{Z}\times\mathbf{Z}\ \to\ \{0,1\}$$

taking (m,n) to 1 if $m = n$ and to 0 otherwise. If q is the nontrivial partition of $\{0,1\}$, then the coarsest partition of $\mathbf{Z}$ compatible with the first argument of α and with the partition q is the partition into infinitely many singletons.

On the other hand, the next Lemma gives a positive result in the context we are interested in.

Lemma 12. *Suppose that in* (10), *all* $X^{(m)}$ *and* Y *are Stone spaces,* α *is continuous, and* q *is a partition of* Y *into open-closed subsets. Then for each* m_0, *the coarsest partition of* $X^{(m_0)}$ *that is compatible with the* m_0*th argument of* α *and with the partition* q *of* Y *is also a partition into open-closed subsets.*

(*Actually, this does not require that all* $X^{(m)}$ *and* Y *be Stone, but only that all* $X^{(m)}$ *other than* $X^{(m_0)}$ *be compact; we have merely stated it under hypotheses we will later find convenient.*)

Proof. First let us simplify the problem: Let us regard the product of all the $X^{(m)}$ other than $X^{(m_0)}$ as a single compact topological space W and let us write $X^{(m_0)}$ as X; thus we have a continuous map $\alpha\colon W\times X \to Y$ and a partition q of Y into open-closed sets. Now let us fix a point $x\in X$, and note that for each $w\in W$, continuity of α implies that there exist open neighborhoods U_w of w and V_w of x such that $\alpha(U_w\times V_w)$ is contained in the open-closed member of q containing $\alpha(w,x)$. By compactness, W is covered by a finite subfamily of the open sets V_w; if we form the intersection of the corresponding finite family of open neighborhoods U_w of x, we get a neighborhood of x which, it is easy to verify, is contained in a single member of the partition of X we are interested in. Since x was an arbitrary point of X, this shows that this partition of X is into *open* sets. But a partition of any

topological space into open sets is a partition into open-closed sets, since the complement of any member U of the partition is a union of open sets, hence open, hence U is closed. □

We now come to our final hurdle. Suppose X is a Stone space given with one or more continuous operations. Given a partition q of X into open-closed subsets, we may form, for each argument of each operation, the coarsest partition compatible with that argument and q. Taking the intersection (coarsest common refinement) of all these partitions and the original partition q, we get a partition q' of X into open-closed subsets, the coarsest such partition which refines q and is compatible with (all the arguments of) all the operations and q – but this does not say that it is compatible with all the operations and itself! We may repeat the process, getting a coarsest partition q'' which refines q', and is compatible with all the operations and q'. If we continue this process recursively, and take the intersection of the resulting chain of partitions, this will be the coarsest partition on X which refines q and is self-compatible under our operations. This is clearly a partition into *closed* subsets, but unless the process stopped after finitely many steps, there will be infinitely many of these, so they will not all be open; in this case, no partition into open-closed subsets refining q can be self-compatible.

Let us note briefly two examples of structures – other than groups – for which this difficulty actually occurs; which, in fact, admit *no* nontrivial partitions into open-closed subsets self-compatible under their operations. (I will show more details, if you want, in class, where I can draw pictures on the board.)

First, let X be a subset of the real unit interval consisting of 0, 1, and the members of a sequence $... < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < ...$ with infimum 0 and supremum 1. This is a Stone space, and the unary operation which takes each x_n to x_{n+1} and fixes 0 and 1 is continuous; but there is no partition of X into open-closed sets self-compatible under this unary operation.

For the second example, form the compact space $2^{\mathbf{Z}}$ of all subsets of $\mathbf{Z}$, and let X be the closed subset consisting of all $x \in 2^{\mathbf{Z}}$ with the property

$$(\forall\, n \in \mathbf{Z}) \quad n-1 \in x,\ n+1 \in x \ \Rightarrow\ n \in x.$$

One finds that this set admits binary operations of greatest lower bound (namely, intersection) and least upper bound (namely, taking the union and then filling in all gaps of length 1). These operations are continuous in the given topology, so X is a topological lattice. But one can show that no nontrivial partition of X into open-closed subsets is self-compatible with respect to both lattice operations.

However, when we consider groups, the stars are favorable. The next result does not refer to topologies, but when combined with the preceding Lemma, it will give us what we have been seeking.

Lemma 13. *Let G be a group, and q a partition of the underlying set of G. Let $\alpha\colon G\times G\times G \to G$ denote the 3-fold multiplication map, $(x,y,z) \mapsto xyz$. Let p be the coarsest partition of G compatible with the middle argument of α and the partition q of G.*

Then p is self-compatible under the operations of G and is a refinement of q, hence is the coarsest partition with these properties.

Proof. Let us first show that p is self-compatible with respect to the first argument of the (binary) group multiplication. That is, assuming y and y' are in the same member of p, we shall prove yu and $y'u$ are also, for any $u \in G$. By definition, the latter condition holds if and only if, for arbitrary x and z, the elements $x(yu)z$ and $x(y'u)z$ are in the same member of q. But by associativity of G, these can be rewritten $x(y)(uz)$ and $x(y')(uz)$, and the assumption that y and y' are in the same member of p says that these 3-fold products are indeed in the same member of q, as required. The proof for the second argument of multiplication is, of course, the same.

For the case of the unary inverse operation, we must show, assuming y and y' in the same member of p, that y^{-1} and y'^{-1} are also, i.e., that for any x and z, the elements $x(y^{-1})z$ and $x(y'^{-1})z$ are in the same member of q. But we can write these two elements as $(xy^{-1})(y')(y'^{-1}z)$ and $(xy^{-1})(y)(y'^{-1}z)$, and again these lie in the same member of q by definition of p. (Alternatively, we could get this result on the inverse operation by noting that the results of the preceding paragraph imply that the set p inherits a multiplication, under which the map from G is a homomorphism of semigroups, and then calling on the well known fact that the image of a group under a semigroup homomorphism inherits an inverse operation.) That p refines q is seen by taking $x = z = 1$. □

A topological group or other algebraic structure which is an inverse limit of finite structures is called *profinite*. (The ''pro'' stands for ''projective'', because inverse limits are sometimes call *projective limits*. In this usage, direct limits are called *inductive limits*. What I am here loosely calling an ''algebraic structure'' can be more precisely defined as an ''algebra in the sense of universal algebra'' [**1**, §1.7 or §8.1].) Clearly, every profinite topological algebraic structure has a Stone space for underlying space. We can now combine the preceding three Lemmas to get the converse in the case of groups:

Theorem 14. *Any topological group G with underlying space a Stone space is profinite.* (*Hence, by Lemma 4, every such group can be represented as the Galois group of an algebraic field extension.*)

Proof. Given any partition q of G into open-closed subsets, let p be the refinement of q constructed in Lemma 13; by Lemma 12 this is also a partition into open-closed sets. Hence condition (iii) of Lemma 11 is satisfied, which implies the desired condition (i) of that Lemma. □

Let us record a couple of easy characterizations of profinite groups, which can be proved independently of the above hard result.

Lemma 15. *For a topological group G, the following conditions are equivalent:*

(i) *G is profinite.*

(ii) *G is compact, and has a neighborhood basis of the identity consisting of open* (*equivalently, open-closed*) *subgroups.*

(iii) *G is compact, and has a neighborhood basis of the identity consisting of open* (*equivalently, open-closed*) *normal subgroups.*

Sketch of Proof. To get the equivalences referred to parenthetically in (ii) and (iii), note that if a subgroup H of a topological group G is open, then so are all the proper right cosets of H, hence H, as the complement of the union of these cosets, is closed. Note also that if H is open and G is compact, then since G is covered by the right cosets of H, it is covered by finitely many, so H has finite index in G.

Now in any topological group G, the open-closed normal subgroups are precisely the kernels of continuous homomorphisms to discrete groups. With these observations (and, again, the idea of the last part of the proof of Theorem 8), the equivalence of (i) and (iii) is straightforward. Clearly, (iii)⇒(ii). To get the converse, note that an open subgroup H, since it has finite index, has only finitely many conjugates, so their intersection is an open normal subgroup contained in H. □

(The equivalence of (ii) and (iii) fails for non-compact topological groups. E.g., the semidirect product of the compact group $(\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}$ and the discrete group $\mathbf{Z}$ acting by the shift has a neighborhood basis of open-closed subgroups, given by any such basis of $(\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}$, but has no open-closed normal subgroups

smaller than $(\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}$ itself.)

5. Remarks on other algebraic structures. We have noted that the statement corresponding to Theorem 14 fails for sets with a single unary operation, and also for lattices. Some examples of algebraic systems for which it *holds*, other than groups, are monoids, semigroups, and associative rings. For monoids, the proof is exactly the same as for groups, using the same operation $(x,y,z) \mapsto xyz$, and merely omitting the discussion of the inverse operation. For semigroups not necessarily having identity element, one can get the result by taking p to be the coarsest partition that is simultaneously compatible with q and with the argument y in each of the four operations $(x,y,z) \mapsto xyz$, $(x,y) \mapsto xy$, $(y,z) \mapsto yz$ and $y \mapsto y$; or, more concisely, by adjoining a topologically isolated identity element, applying the result for monoids, and then discarding the identity element. For associative rings with 1, one uses the argument y of the 4-ary operation $(w,x,y,z) \mapsto w + xyz$; the reader is invited to work this case out, and see how nicely the ring identities yield the conditions needed. For nonunital associative rings, one can replace this with a family of operations, as for nonunital semigroups.

Note that the operations $\wedge$ and $\vee$ of a lattice are both semigroup operations. So by the above observations, a topological lattice with a Stone topology is profinite with respect to each of its operations separately (i.e., as an upper and as a lower semilattice), but not, in general, simultaneously! (For the reader who knows a bit more lattice theory, I will mention that a negative example can also be given for *modular* lattices, but that *distributive* lattices with Stone topologies are profinite, as one shows using the argument y of the operation $x \wedge (y \vee z)$.)

There is a parallel set of questions that can be treated in the same way as those treated here: If k is any field, then an inverse limit V of finite-dimensional discrete k-vector spaces, though not compact unless k is finite, will have a property called *linear compactness*: any family of *cosets of closed k-subspaces* of V which has the finite-intersection property has nonempty intersection. The topology on such an inverse limit V is also clearly what is known as a *linear* topology, meaning that open vector subspaces of V form a neighborhood basis of zero. One can show that the linearly compact Hausdorff linearly topologized vector spaces are precisely the inverse limits of finite dimensional discrete vector spaces, and establish a result analogous to Theorem 14, namely that an associative algebra with such a topology is an inverse limit of finite-dimensional discrete associative algebras. On the other hand, there are simple examples showing that the corresponding statement fails for other structures, such as Lie algebras.

REFERENCES

1. George M. Bergman, *An Invitation to General Algebra and Universal Constructions*, Berkeley Mathematics Lecture Notes, vol. 7, U. C. Berkeley Department of Mathematics, 1995.

2. Thomas W. Hungerford, *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, 1974.

3. John L. Kelley, *General Topology*, University Series in Higher Mathematics, van Nostrand, 1955; Graduate Texts in Mathematics, vol. 27, Springer-Verlag, 1975. (My copy is the van Nostrand edition. I hope that the references I give are correct for the Springer edition as well.)

4. Serge Lang, *Algebra*, 3rd ed., Addison-Wesley, 1993.

5. Saunders Mac Lane, *Categories for the Working Mathematician*, Graduate Texts in Math., vol. 5, Springer-Verlag, 1971.

Notes on composition of maps

1. Some notational conventions. There are four common notations for the image of an element x under a function f:

(1) Left multiplicative notation: fx.

(2) Standard functional notation: $f(x)$.

(3) Right multiplicative notation: xf.

(4) Exponential notation: x^f.

Consider the following rule of thumb:

Convention 1. *Suppose we are given functions* $X \xrightarrow{f} Y \xrightarrow{g} Z$. *Then if* f *and* g *belong to a class of functions written on the* left *of their arguments* (*as in notation* (1) *or* (2) *above*), *write their composite as* gf *or* $g \circ f$. *If, on the other hand,* f *and* g *belong to a class of functions written on the* right *of their arguments* (*as in* (3) *or* (4)), *call their composite* fg *or* $f \circ g$.

Recall that whichever order of composition one uses, composition of functions is associative. The advantage of Convention 1 is that it creates, by definition, a ''second associative law'' for functions and their arguments. That is, in notations (1) and (3),

(5) $$(fg)x = f(gx), \qquad \text{respectively} \qquad x(gf) = (xg)f.$$

In either case, this associativity makes it possible to drop parentheses, and write fgx or xgf.

Essentially all writers follow Convention 1 in a ''weak sense''; namely, if they write most functions to the left of their arguments, they write composites of such functions in the ''lefthand'' manner, while if they write most functions to the right of their arguments, they use ''righthand'' composition. However, most authors try to keep all function symbols on the same side of their arguments, and if they occasionally give in to special circumstances and write functions on the other side, they often stick to their original definition of composition, and thus have ''scrambled associative laws'' such as $(fg)x = g(f(x))$ holding in these situations.

I claim, however, that in many situations, especially linear algebra (i.e., module theory), it is useful to work both with functions written on the left *and* and with functions written on the right, and to compose each type of function according to the appropriate rule.

Let us make clear the price. First, we have to apply contradictory definitions of composition to different functions occurring in the same discussion. As a consequence, we cannot uniformly make statements like ''a left-invertible function is one-to-one''. Rather, this statement applies to classes of functions that we write on the left of their arguments, while left-invertible members of classes of functions written on the right of their arguments are *surjective*. However, since in reading different authors, we have to deal sometimes with functions written on one side and sometimes with functions written on the other, it is not that hard to adjust to different notations being used within one text, as long as the motivation for the difference is made clear.

Secondly, Convention 1 gives us no way of denoting the composite of a function f, g etc. written on the left with another function a, b etc. written on the right. It is easy to denote the action of such a composite on an element x, e.g., $(fx)a$ or $f(xa)$; but if we need a symbol for the composite itself, we must resort to some artifice; e.g., we might introduce new symbols $\tilde{a}$, $\tilde{b}$, etc., denoting the same functions as a, b etc., but written to the left of their arguments. We can then write composites $\tilde{a}f$, $f\tilde{a}$ etc.,

without violating our Convention. (Note that if we do this, then though $\tilde{a}$, $\tilde{b}$, $\widetilde{ab}$ name the same functions as a, b, ab, we have $\widetilde{ab} = \tilde{b}\tilde{a}$.)

But why would we want to write some maps to the left of their arguments, and others to the right? Because of another very useful principle:

Convention 2. *If one class of functions is characterized as commuting with* (''*respecting*'') *the functions in another class, write members of the two classes on opposite sides of their arguments.*

Thus, if f belongs to a class of functions written on the left, and a to a class of functions respecting members of that class, and written on the right, then the fact that f and a respect one another becomes a third ''associative law'':

$$f(xa) = (fx)a, \tag{6}$$

(again allowing us to drop parentheses and write fxa if we wish).

Our final, and perhaps most important Convention concerns the situation where we have not merely a family of maps on a set, but an abstract structure of some sort ''acting on'' a set. (For instance, to anticipate the subject of the next section, if M is a module over a ring R, then the elements of R act on the underlying additive group of M.) In such situations, there is often a ''multiplication'' operation on the object that is acting, and we apply

Convention 3. *Suppose a set R has an operation denoted multiplicatively, $(f,g) \mapsto fg$, and we have an ''action'' of R on a set X, i.e., a map $R \times X \to X$. Then write this action as left or right ''multiplication'' if the action on X of a ''product'' of elements in R is given by the composite of the actions of the respective elements, in the order specified by Convention 1 for operations written on that side.*

The point of this Convention is clear. Let us now examine a particularly important class of applications.

2. Modules (and bimodules). Suppose a ring R acts on an additive group M in such a way that on applying to an element $x \in M$ an element $r \in R$, and then applying to the result another element $s \in R$, the final result is the same as that of applying $rs \in R$ to x. Then by Convention 3, it is natural to denote this action by writing elements of R on the right of elements of M. Assuming that the action is bilinear and that the identity element of R acts as the identity endomorphism of M, the resulting structure on M is that of a *right R-module*. If, instead, R acts so that the action of r followed by that of s is equivalent to the action of sr, then we write elements of R on the left of elements of M, and M becomes a *left R-module*. (Incidentally, one sometimes signals that an object M is a right or left R-module by writing it M_R, respectively ${}_R M$.)

Thus the essential difference between right and left modules is in the way multiplication of elements of R relates to composition of the associated operations on M. The difference in notation (xr vs. rx) is merely a corollary of this, though it gives the names to these two types of modules.

If M is a *right* R-module, then by Convention 2, a system of R-module *endomorphisms* of M should be written on the *left* of M. Such a system, if it is closed under composition and addition and contains the identity endomorphism, will form a ring. If a ring S acts on the left of a right R-module M by such module endomorphisms – or expressed more symmetrically, if a ring S acts on the left, and a ring R acts on the right, on the same abelian group M, and these actions respect one another – then the resulting structure is what is called an (S,R)-*bimodule* ${}_S M_R$. A familiar example is when M is the set ${}^m R^n$ of

all $m \times n$ matrices over a ring R; this can easily be seen to form an $({}^mR^m, {}^nR^n)$-bimodule.

3. Multiplication and composition. The ''additional associativity laws'' (5) and (6) to which our conventions lead can be of more than just aesthetic and mnemonic value. A common way of constructing models of mathematical structures is as sets of *functions* – rings of operators, vector spaces of continuous functions on a topological space, etc.. In such cases, one often defines the operations of these objects by *composition* of functions. Since composition of functions is associative, we can only hope to realize a structure by such a model if it satisfies an appropriate associative law. Of course, if a structure satisfies a scrambled associative law such as $(f * g)x = g(fx)$, it may be possible to realize it by a system of functions with operation $*$ defined by $f * g = g \circ f$. But it is clearly easiest to see our way if we write things so that true associativity holds.

Since matrices are a model of linear maps among free modules, the matrix example discussed above is a model of the definition of bimodule, based on functional composition. Here is a related class of examples.

Suppose R is a ring, and M and N are right R-modules. Recall that the set of right R-module homomorphisms $M \to N$ forms an additive group, but (for R not commutative), does not in general form an R-module. (If we attempt to multiply a homomorphism h by an element $r \in R$, the result is not in general an R-module homomorphism, because multiplication by r itself is not; i.e., r does not in general commute with other elements of R.) However, if N is not only a right R-module but an (S, R)-bimodule for some ring S, then we can left-compose R-module homomorphisms $M \to N$ with the actions of elements of S (which, we have seen, are R-module endomorphisms of N) to get new R-module homomorphisms $M \to N$. In this way, the set of these right R-module homomorphisms becomes a left S-module. If, instead, we are given a (T, R)-bimodule structure, on M we can compose R-module homomorphisms $M \to N$ on the right with actions of elements of T, and our hom-set becomes a right T-module. If we have both kinds of structures, then our hom-set, $\mathrm{Hom}_R({}_TM_R, {}_SN_R)$, becomes an (S, T)-bimodule.

In particular, if M is a right R-module, and we form the set $\mathrm{Hom}_R(M, R)$, then since R is an (R, R)-bimodule, its left module structure induces a structure of left R-module on this set; this left R-module is called the *dual* of the right R-module M. Likewise, the dual of a *left* R-module, i.e., the set of left linear functionals on it, has a natural structure of *right* R-module.

This duality gives a contravariant functor $^\vee$ from right R-modules to left R-modules, and vice versa. If one writes all morphisms on the same side of their arguments, this contravariance takes the form $(fg)^\vee = g^\vee f^\vee$. However, if we adjust the sides on which morphisms are written according to Convention 2, and compose them accordingly, we get $(fg)^\vee = f^\vee g^\vee$; though we must still regard $^\vee$ as contravariant! If there is no danger of ambiguity, we can even drop the symbol ''$^\vee$'' from our morphisms, applying a homomorphism $f: M \to N$ to the left of elements of M to get elements of N, but to the right of elements of $N^\vee$ to get (by composition) elements of $M^\vee$.

It is not hard to see that the dual of a free right R-module of finite rank is a free left R-module of the same rank. If we regard the set ${}^mR^n$ of all $m \times n$ matrices as representing the set of module homomorphisms from the standard free right R-module on n generators to the standard free right R-module on m generators, then the above observation shows that the free left R-module on n generators can be identified with the $m = 1$ case of this hom-set, i.e., the space row vectors of length n. Clearly, the free right R-module on m generators can itself be identified with the $n = 1$ case, i.e., the space column vectors of height m. Thus, the free modules have been incorporated right into the system of matrices designed to represent their homomorphisms, and this system is symmetric, with free left modules and free right modules now in completely analogous roles. Suppressing superscripts ''^{1}'', this leads to the

notation ^mR for the free right R-module on m generators, and R^m for the corresponding free left R-module, often used by ring-theorists.

Matrices also illustrate strikingly our earlier observation that if we write some maps on the right and some on the left, the properties of ''left'' and ''right'' invertibility will depend on the sort of morphisms we are looking at. A left-invertible $m \times n$ matrix represents both a *surjective* linear map from column vectors of height n to column vectors of height m, and a *one-to-one* linear map from row vectors of length m to row vectors of length n.

4. Assorted remarks.

4.1. Superscript notation is most frequently used for automorphisms. In particular, the inner automorphisms of a group are often written $h^{-1}gh = g^h$. This notation is consistent with Convention 3, since $(g^h)^{h'} = g^{hh'}$, while if we defined g^h to mean hgh^{-1} (also often called the ''conjugate of g by h''), this would fail. A way to remember the correct choice is to note that g^h should have the form ''$...gh$'', so that when we compose two such operations we will get $(g^h)^{h'} = ...ghh' = g^{hh'}$; if we defined the conjugate of g by h to have the form ''$hg...$'' it would, instead, behave like a left-multiplication operator. Group-theorists sometimes do use this definition, and write this conjugate ^hg.

4.2. Suppose we are given an abelian group M, and an additive group R of endomorphisms of M closed under composition, and containing the identity endomorphism. It is amusing to note that until we decide on the notation we are going to use – writing these endomorphisms on the left and composing them accordingly, or writing them on the right and composing them accordingly – we can't formally say what their ring structure is, and whether M is a right or left module over this ring.

Similarly, if in a paper one makes a brief remark about functions in a context where it is not completely clear whether these would be written on the right or on the left of their arguments, there can be danger of ambiguity if one says something like ''composing f on the right with an appropriate automorphism α, ...''. I sometimes avoid this problem by saying ''precomposing f ...'', i.e., I use ''precompose'' and ''postcompose'' to mean, respectively, that the other map is to act before or after f.

4.3. If M is a right or left R-module, then its double dual $M^{\vee\vee}$ is a module of the same kind, and, just as in elementary linear algebra, there is an obvious natural module homomorphism $M \to M^{\vee\vee}$. If this map is an isomorphism, M is called *reflexive*. We see from previous remarks that free modules of finite rank are reflexive.

We commented in the preceding section that in writing the dual $f^\vee$ of a module homomorphism $f: M \to N$, one might drop the superscript $^\vee$ if there is no danger of ambiguity. The case where this is so is essentially that in which M and N are reflexive. In that case, it is easy to show that the functor $^\vee$ gives a bijection between right module homomorphisms $M \to N$ and left module homomorphisms $N^\vee \to M^\vee$; in other cases, the map in question may fail to be one-to-one and/or onto, and one needs to distinguish between the two hom-sets. It is the fact that free modules of finite rank *are* reflexive that allows us to use matrices both for maps among free left modules and maps among free right modules, in such a carefree fashion.

4.4. Here is an example of considerations like those of Convention 3, but in an important *nonassociative* context. A *Lie algebra* L over a field k is a k-vector space given with a bilinear operation called the ''Lie bracket'', not in general associative, and not written ab but $[a, b]$, which satisfies identities modeled on the properties of the *commutator* operation $[a, b] = ab - ba$ in an associative k-algebra. A ''module'' over the Lie algebra L is a vector space M given with a ''multiplication'' $(a, x) \mapsto ax$

($a \in L$, $x \in M$). Here one does not require the result of multiplying an element x by a bracket $[a, b]$ to coincide with the result of applying a and b to x in either of the two orders, $a(bx)$ or $b(ax)$. Instead, one requires that $[a, b]x = a(bx) - b(ax)$; this turns out to be the ''right'' definition for the subject.

4.5. One objection that could be raised to the idea of using two different notations for composition of maps on the same objects is that it would make for great notational confusion in considering the *category* of those objects. However, note that if we are considering, say, right modules over a ring R, and writing their homomorphisms on the left, following Convention 2, it is only these homomorphisms, and not the actions of elements of R, that occur as morphisms in the module category; so in considering that category, our Convention causes no problem. It is only if we want to regard these modules as objects in the category of abelian groups, and work both with the module homomorphisms and with the endomorphisms arising from the action of R, that we get a conflict. In that situation, it is probably best to set Convention 2 aside; though it is conceivable that if we wanted to work extensively with this sort of situation in a general category $\mathbf{C}$, we could set up conventions that tell us to write some morphisms in $\mathbf{C}$, and others in the opposite category, $\mathbf{C}^{\mathrm{op}}$.

4.6. A similar problem arises with homomorphisms of *bimodules*. If M, N are (S, T)-bimodules, and $\varphi\colon M \to N$ a bimodule homomorphism, i.e., both an S-module and a T-module homomorphism, then since it commutes with both the action of S on the left and the action of T on the right, Convention 2 gives us contradictory instructions as to where to write it. Here, obviously, we cannot fully follow that Convention! (If our writing system did not encourage linearity, we might write φ above or below the element to which it is to be applied. The condition that our homomorphism respect the actions of both S and T would again allow us to drop ''parentheses'' in such expressions; but otherwise the idea seems to have little to say for it.)

In this context I will mention that ring-theorists sometimes find it useful to note that an (S, T)-bimodule is equivalent to an abelian group with *left* actions of the rings S and T^{op} which commute with one another elementwise; equivalently, to a left module over the ring $S \otimes T^{\mathrm{op}}$, and they investigate properties of the latter ring as a tool in studying such bimodules.

5. Commutative rings, and rings with involution. The opposite to the sort of problem described in the last remark arises when we consider modules over a commutative ring R, since there is no distinction between saying that multiplication by $r \in R$ followed by multiplication by $s \in R$ acts like multiplication by rs, and saying that it acts like multiplication by sr.[1] We can only speak of objects on which this composite behaves like the common value $rs = sr$, which are, of course, called *R-modules*. Here we can decide arbitrarily to write the actions of elements of R on one side, say the right, and homomorphisms on the other; in particular, we can decide arbitrarily to identify a free R-module M of rank n with the standard left module of column vectors of height n; in that case the dual module (often called M^*, but called by Lang $M^\vee$) will be identified with the standard left module of row vectors. Although modules of row and of column vectors are isomorphic, if they arise differently in a given context, this formal distinction can actually help us keep track of things; e.g., distinguish modules that arise covariantly or contravariantly from given data.

An important context where we *cannot* maintain such a distinction is when we are given a free module

[1] And the Lord said, '... should I not have pity on Nineveh, that great city, wherein are more than six score thousand persons that cannot discern between their right hand and their left hand, and also much cattle?' Jonah 4:10-11.

M over a commutative ring, together with an isomorphism with its dual. This is equivalent to a nonsingular bilinear map $M \times M \to R$. (An inner product on a real vector space is an example of such a map.) As an example of the difficulty and its consequences, consider a linear map $f: M \to N$ of R-modules. According to our Conventions, if we treat M and N as right R-modules, we want to write f to the left of its argument; and the dual map $N^\vee \to M^\vee$, given by composition with f, can be denoted by writing f to the right of members of $N^\vee$. But if M and N are given with isomorphisms that ''identify'' them with $M^\vee$ and $N^\vee$, we want to write homomorphisms between the latter modules on the left.

In fact, if we take the above induced map $N^\vee \to M^\vee$ and regard it as a map $N \to M$, the result is what is called the *transpose* ${}^t f$ of f; and the consequence of our switching it from the right to the left of its argument is the familiar equation ${}^t(fg) = {}^t g\, {}^t f$.

Some of these considerations on commutative rings can be generalized to not necessarily commutative rings R given with an *involution*, that is, a map $*: R \to R$ satisfying the identity $r** = r$, which respects addition and *reverses* multiplication: $(rs)* = s*r*$. If R is a ring with involution, then any right R-module M can be made a left R-module by defining, for $r \in R$, $x \in M$,

$$rx = xr*.$$

Thus the dual of a left module can be turned back into a left module, so that we can again consider left modules given with isomorphisms with their duals, and again use duality to get, for every map $M \to N$ between modules given with such isomorphisms, a ''transpose'' map $N \to M$. In the case of standard free modules, made isomorphic with their duals by taking each member of the standard basis to the corresponding member of the standard dual basis, this ''transpose'' operation corresponds to transposing the matrix representing a module homomorphism, and applying $*$ to each entry.

Although I have brought this up as a way to extend certain ideas of commutative ring theory to the noncommutative case, the most familiar example is actually one in which R is a commutative ring, namely the complex numbers, and $*$ is complex conjugation. Then the above considerations yield the familiar ''conjugate transpose'' operation on matrices.

For the reader curious to see some noncommutative rings with involution, we can begin by biting our tail, and giving the example ${}^n R^n$, where R is any commutative ring, and $*$ is the transpose operation; a variant is ${}^n\mathbf{C}^n$ with $*$ the conjugate transpose operation. A somewhat related example is the division ring of quaternions, on which one can easily verify that $a + bi + cj + dk \;\mapsto\; a - bi - cj - dk$ is an involution. For a couple of examples a little farther afield, let k be any field, and form the group algebra kG of any group, or the free associative algebra $k\langle X\rangle$ on any set. In the former case, the map that is the identity on k and sends each $g \in G$ to g^{-1} is an involution; in the latter, the map that is the identity on k and reverses the order of factors in every monomial is one.

6. A final word. Even when no formal difficulties arise, I do not *always* follow, nor recommend that anyone *always* follow the system of Conventions given in §1 above. Sometimes one wishes to use ''standard functional notation'' $f(x)$ to make very clear what is function and what is argument, even if this leads to violations of Conventions 2 and 3; sometimes there are other reasons.

But in many situations, those Conventions are of great value, especially in understanding the structures of hom-sets of modules, and in working with matrices, so I urge the student to familiarize himself or herself with them.

Tensor algebras, exterior algebras, and symmetric algebras

1. Basics. Throughout this note, k will be a fixed commutative ring.

Recall that if M and P are k-modules, a k-multilinear map $\mu: M^n \to P$ is called *alternating* if $\mu(x_1, \dots, x_n) = 0$ whenever two successive arguments x_i, x_{i+1} are equal. From this condition one easily shows that transposing two successive arguments reverses the sign of $\mu(x_1, \dots, x_n) = 0$, from which it follows that any permutation of the arguments multiplies the value by the sign of that permutation; hence, from the original condition we see that the value is zero if *any* two terms are equal.

We recall likewise that a k-multilinear map $M^n \to P$ is called *symmetric* if it is unchanged under transposing successive arguments; equivalently, if it is unchanged under all permutations of the arguments.

Let us now define some universal maps.

Definition 1. *Let M be a k-module.*

By $\otimes^n M$ we shall denote the k-module with a universal k-multilinear *map of M^n into it, written $(x_1, \dots, x_n) \mapsto x_1 \otimes \dots \otimes x_n$. This module is called the n-fold* tensor power *of M.*

By $\Lambda^n M$ we shall denote the k-module with a universal alternating *k-multilinear map of M^n into it, written $(x_1, \dots, x_n) \mapsto x_1 \wedge \dots \wedge x_n$. This module is called the n-fold* exterior power *of M.*

By $S^n M$ we shall denote the k-module with a universal symmetric *k-multilinear map of M^n into it, written $(x_1, \dots, x_n) \mapsto x_1 \dots x_n$. This module is called the n-fold* symmetric power *of M.*

Each of these modules can clearly be constructed directly by generators and relations. $\otimes^n M$ can also be constructed as $M \otimes_k (M \otimes_k (\dots))$ (with n M's), while $\Lambda^n M$ and $S^n M$ can be constructed as factor-modules of $\otimes^n M$ by the submodules generated by all elements of the forms

$$x_1 \otimes \dots \otimes y \otimes y \otimes \dots \otimes x_n,$$

respectively

$$x_1 \otimes \dots \otimes y \otimes z \otimes \dots \otimes x_n \ - \ x_1 \otimes \dots \otimes z \otimes y \otimes \dots \otimes x_n.$$

(where each term is $x_1 \otimes \dots \otimes x_n$ modified only in two successive places).

Let us make the convention that $\otimes^1 M$, $\Lambda^1 M$ and $S^1 M$ are all identified with M, while $\otimes^0 M$, $\Lambda^0 M$ and $S^0 M$ are all identified with k.

The universal properties defining our three constructions can be conveniently restated as follows:

A k-module homomorphism with domain $\otimes^n M$ is uniquely determined by specifying it on the elements $x_1 \otimes \dots \otimes x_n$ for all $x_1, \dots, x_n \in M$, in such a way that the value is a *k-multilinear* function of $x_1, \dots, x_n$.

A k-module homomorphism with domain $\Lambda^n M$ is uniquely determined by specifying it on the elements $x_1 \wedge \dots \wedge x_n$ for all $x_1, \dots, x_n \in M$, so that the value is an *alternating* k-multilinear function of $x_1, \dots, x_n$.

A k-module homomorphism with domain $S^n M$ is uniquely determined by specifying it on the elements $x_1 \dots x_n$ for all $x_1, \dots, x_n \in M$, so that the value is a *symmetric* k-multilinear function of $x_1, \dots, x_n$.

Each of $\otimes^n$, Λ^n, S^n may be made into a functor. Given $f: M \to N$, one defines, for example, $\Lambda^n f: \Lambda^n M \to \Lambda^n N$ as the unique module homomorphism taking $x_1 \wedge \dots \wedge x_n$ to $f(x_1) \wedge \dots \wedge f(x_n)$ for all $x_1, \dots, x_n \in M$. This homomorphism exists because the latter expression in $x_1, \dots, x_n$ is alternating and

k-multilinear.

Let us now show how to multiply a member of $\otimes^m M$ by a member of $\otimes^n M$ to get a member of $\otimes^{m+n} M$, and similarly for the other constructions. The universal properties of our module constructions characterize the module homomorphisms, i.e. *linear* maps, definable on them, but we now want to construct *bilinear* maps. The next Lemma allows us to translate existence and uniqueness results for the former into existence and uniqueness results for the latter. (Note that in that Lemma, the subscripts i and j index generators, while i' and j' index relations.)

Lemma 2. *Let A be a k-module presented by generators X_i $(i \in I)$ and relations $\Sigma_i \, \alpha_{i'i} X_i = 0$ $(i' \in I')$, and B a k-module presented by generators Y_j $(j \in J)$ and relations $\Sigma_j \, \beta_{j'j} Y_j = 0$ $(j' \in J')$.*

Then for any k-module C, a k-bilinear map $A \times B \to C$ is uniquely determined by specifying the image Z_{ij} of each pair of generators (X_i, Y_j), in such a way that for each $j \in J$, $i' \in I'$ one has $\Sigma_i \, \alpha_{i'i} Z_{ij} = 0$, and for each $i \in I$, $j' \in J'$ one has $\Sigma_j \, \beta_{j'j} Z_{ij} = 0$.

(*Equivalently, the k-module $A \otimes_k B$ may be presented by the generators $X_i \otimes Y_j$ and the two sets of relations $\Sigma_i \, \alpha_{i'i} X_i \otimes Y_j = 0$ $(i' \in I', j \in J)$ and $\Sigma_j \, \beta_{j'j} X_i \otimes Y_j = 0$ $(j' \in J', i \in I)$.*)

Proof. Exercise. □

We deduce

Lemma 3. *Let M be a k-module and m, n nonnegative integers. Then there exist*
a unique k-bilinear map $\otimes^m M \times \otimes^n M \to \otimes^{m+n} M$ which carries each pair

$$(x_1 \otimes \ldots \otimes x_m, \ x_{m+1} \otimes \ldots \otimes x_{m+n}) \ \text{ to } \ x_1 \otimes \ldots \otimes x_{m+n},$$

a unique k-bilinear map $\Lambda^m M \times \Lambda^n M \to \Lambda^{m+n} M$ which carries each pair

$$(x_1 \wedge \ldots \wedge x_m, \ x_{m+1} \wedge \ldots \wedge x_{m+n}) \ \text{ to } \ x_1 \wedge \ldots \wedge x_{m+n}, \ \text{ and}$$

a unique k-bilinear map $S^m M \times S^n M \to S^{m+n} M$ which carries each pair

$$(x_1 \ldots x_m, \ x_{m+1} \ldots x_{m+n}) \ \text{ to } \ x_1 \ldots x_{m+n}.$$

Proof. We are describing these maps in terms of their actions on pairs of generators of the two modules in question, so by Lemma 2, it suffices to verify that these assignments satisfy the equations corresponding to the defining relations for the two modules. These equations are equivalent to saying that the assignments should be multilinear (respectively alternating multilinear, respectively symmetric multilinear) in $x_1, \ldots, x_m$, and also in $x_{m+1}, \ldots, x_{m+n}$, which is clearly true in each case. □

Let us now form the k-modules

$$\begin{aligned} k\langle M \rangle &= \oplus_n \, (\otimes^n M), \\ \Lambda M &= \oplus_n \, (\Lambda^n M), \\ S M &= \oplus_n \, (S^n M). \end{aligned}$$

We define a bilinear operation on each of them by letting it act on pairs of elements from the mth and nth summands by the operations described in Lemma 3, and extending it to sums of elements from different summands by bilinearity. We can now assert

Theorem 4. *The multiplications on* $k\langle M\rangle$, ΛM *and* SM *defined above are structures of* (*associative unital*) *k-algebra.*

The algebra $k\langle M\rangle$, *called the* tensor algebra *on* M, *is universal among k-algebras given with k-module homomorphisms of* M *into them.*

The algebra ΛM, *called the* exterior algebra *on* M, *is universal among k-algebras given with k-module homomorphisms of* M *into them such that the images of all elements of* M *have zero square.*

The algebra SM, *called the* symmetric algebra *on* M, *is universal among k-algebras given with k-module homomorphisms of* M *into them such that the images of elements of* M *commute with one another, and is also universal among all* commutative *k*-algebras *given with k-module homomorphisms of* M *into them.*

Proof. To verify associativity of an operation known to be *k*-bilinear, it suffices to verify it on 3-tuples of elements from any set of *k*-module generators. Thus, for instance, associativity of the multiplication of ΛM can be verified by checking it on 3-tuples $x_1 \wedge \ldots \wedge x_m$, $x_{m+1} \wedge \ldots \wedge x_{m+n}$, $x_{m+n+1} \wedge \ldots \wedge x_{m+n+p}$. But it is immediate from the definition that the product, formed either way, is $x_1 \wedge \ldots \wedge x_{m+n+p}$. The associativity of the other two constructions is proved the same way.

Thus, our three objects are indeed a *k*-algebra with a *k*-module homomorphism of M into it (namely, the inclusion of $\otimes^1 M$), a *k*-algebra with a *k*-module homomorphism of M into it such that the images of all elements of M have square zero, and a *k*-algebra with a *k*-module homomorphism of M into it such that the images of all elements of M commute with one another, and in this last case it follows that the algebra is commutative. To complete the proof of the Theorem, we need to show that these algebras are universal for these properties.

Consider first any *k*-algebra R with a *k*-module homomorphism $f\colon M \to R$. We see that for each n, the map $M^n \to R$ given by $(x_1, \ldots, x_n) \mapsto f(x_1) \ldots f(x_n)$ is *k*-multilinear, hence by the universal property defining $\otimes^n M$, it induces a *k*-module homomorphism $\otimes^n M \to R$. This family of maps yields a module homomorphism $k\langle M\rangle \to R$. When we consider the behavior of this map on the *k*-module generators $x_1 \otimes \ldots \otimes x_n$ of $k\langle M\rangle$, we find that the way we have defined multiplication in this algebra makes the map respect multiplication of these generators. From this and the fact that it is *k*-linear, it follows that it respects multiplication on all elements, establishing the asserted universal property.

Likewise, given a *k*-algebra R with a *k*-module homomorphism $f\colon M \to R$ such that for every $x \in M$, $f(x)^2 = 0$, the map $M^n \to R$ taking $(x_1, \ldots, x_n)$ to $f(x_1) \ldots f(x_n)$ is *k*-multilinear and alternating. Thus it extends to a map on $\Lambda^n M$, and the proof is completed as before.

In the case of the universal property of the symmetric algebra, we similarly see that $(x_1, \ldots, x_n) \mapsto f(x_1) \ldots f(x_n)$ is a *symmetric* *k*-multilinear map, and complete the proof in the same way. From the observation that SM is commutative and the above universal property, it follows that it is also universal among commutative *k*-algebras with maps of M into them. □

It follows, either from the explicit description or by the universal property, that each of the above constructions gives a *functor* from *k*-modules to *k*-algebras.

What do these three sorts of algebras look like? Let us get explicit descriptions in the case where M is a free module.

Theorem 5. *Let* M *be a free k-module on a basis* X. *Then:*

$k\langle M\rangle$ *is the free associative k-algebra* $k\langle X\rangle$, *equivalently, the semigroup algebra on the free semigroup* $\langle X\rangle$. *It has as a k-module basis the set of all products* $x_1 \otimes \ldots \otimes x_n$ (*or, using ordinary*

multiplicative notation, as is common when this ring is regarded as a free algebra or a semigroup algebra, the set of products $x_1 \dots x_n$) *for* $x_1, \dots, x_n \in X$. *Thus, if* X *is a finite set* $\{x_1, \dots, x_r\}$, *then for each* n, $\dim_k(\otimes^n M) = r^n$.

ΛM *may be presented by the generating set* X *and the relations* $x \wedge x = 0$, $x \wedge y + y \wedge x = 0$ $(x, y \in X)$. *If a total ordering ''*$\leq$*'' is chosen on* X, *then a* k*-module basis for* ΛM *is given by those products* $x_1 \wedge \dots \wedge x_n$ *with* $x_1 < \dots < x_n \in X$. *In particular, if* X *is a finite set* $\{x_1, \dots, x_r\}$, *then a basis is given by those products* $x_{i_1} \wedge \dots \wedge x_{i_n}$ *with* $1 \leq i_1 < \dots < i_n \leq r$, *hence for each* n, $\dim_k(\Lambda^n M) = \binom{r}{n}$.

SM *may be presented by the generating set* X, *and relations* $xy = yx$ $(x, y \in X)$, *and is the* (*commutative*) *polynomial algebra* $k[X]$, *equivalently, the free commutative* k*-algebra on* X, *equivalently, the semigroup algebra on the free commutative semigroup on* X. *If a total ordering ''*$\leq$*'' is chosen on* X, *then a* k*-module basis for* SM *is given by those products* $x_1 \dots x_n$ *with* $x_1 \leq \dots \leq x_n \in X$. *If* X *is a finite set* $\{x_1, \dots, x_r\}$, *then the elements of this basis can be written* $x_1^{i_1} \dots x_r^{i_r}$ *with* $i_1, \dots, i_r \geq 0$, *and for each* n, $\dim_k(S^n M) = \binom{r+n-1}{n}$.

Proof. The universal property of $k\langle M\rangle$ combined with the universal property of the free module M on X imply that $k\langle M\rangle$ has the universal property of the free algebra $k\langle X\rangle$; and the semigroup algebra on the free semigroup also has this universal property, hence $k\langle M\rangle$ has for basis the free semigroup on X. The other two algebras can now be obtained by imposing additional relations on this free algebra, and the indicated bases are easily established using the diamond lemma [**1**, §1]. □

Remark: The constructions of $\otimes^n M$ and $k\langle M\rangle$ can be carried out and their universal properties verified in the more general context where k is a not-necessarily-commutative ring and M is a k-bimodule. (In this context, $k\langle M\rangle$ is called the ''tensor k-ring on M''.)

The analogs of the exterior and symmetric algebras can, of course, be formally defined in this context as well, but these turn out to be degenerate, with all symmetric or exterior powers of a module beyond the first reducing to cases of the constructions for modules over a commutative ring developed above. For instance, suppose M is a bimodule over a ring R, and $x, y \in M$, $a \in R$. Then in the ''symmetric square'' of M we see that $axy = ayx = (ay)x = x(ay) = xay$; thus the distinction between ax and xa has been lost in this bimodule. Moreover, if b is another element of R, we see using the above computation that $(ab)xy = x(ab)y = (ax)by = b(ax)y = (ba)xy$, so left multiplication by ab and ba are the same; i.e., the left R-module structure of S^2M is induced by a module structure over the abelianization $R/[R:R]$ of R. From such observations one can show that for $n \geq 2$, $S^n M$ can be identified with the nth symmetric power of the $R/[R:R]$-module obtained by dividing M by the sub-bimodule spanned by all elements $ax - xa$ $(a \in R,\ x \in M)$, and the corresponding observations apply to $\Lambda^n M$.

2. Exterior powers and determinants. Note that if, as in the preceding Theorem, M is a free k-module on r generators $x_1, \dots, x_r$, then $\Lambda^r M$ will be a free k-module on one generator $x_1 \wedge \dots \wedge x_r$. Hence if $f\colon M \to M$ is a module endomorphism, the induced endomorphism $\Lambda^r f\colon \Lambda^r M \to \Lambda^r M$ will take $x_1 \wedge \dots \wedge x_r$ to $d(f)\, x_1 \wedge \dots \wedge x_r$ for some constant $d(f) \in k$. Let us note the properties these constants $d(f)$ must have.

Since specifying f is equivalent to specifying $f(x_1), \dots, f(x_r)$, the element $d(f)$ can be thought of as a function of this r-tuple of elements of M. As such it can be evaluated by forming the product $f(x_1) \wedge \dots \wedge f(x_r)$ in $\Lambda^r M$, and writing this as a multiple of $x_1 \wedge \dots \wedge x_r$; the coefficient is $d(f)$. It is easy to see from this formulation that $d(f)$ is an alternating multilinear functional of $f(x_1), \dots, f(x_r)$,

which takes the identity map to 1. We thus recognize d as the determinant function! In fact, this can be taken as the *definition* of determinant, and the standard properties of that construction derived elegantly from this definition.

One such property is that every alternating r-linear functional on M is a multiple of the determinant. Let us, more generally, prove such a result for alternating r-linear functions from M into *any* k-module. Let $a: M^r \to N$ be such a function. Then by the universal property of $\Lambda^r M$ we can write $a(y_1, \dots, y_r) = \alpha(y_1 \wedge \dots \wedge y_r)$ for some k-module homomorphism $\alpha: \Lambda^n M \to N$. Let $\alpha(x_1 \wedge \dots \wedge x_r) = u$. Take any $y_1, \dots, y_r \in M$, and let f denote the unique endomorphism of M that carries the elements $x_1, \dots, x_r$ to $y_1, \dots, y_r$ respectively. Then $a(y_1, \dots, y_r) = a(f(x_1), \dots, f(x_r)) = \alpha(f(x_1) \wedge \dots \wedge f(x_r)) = \alpha(d(f)\, x_1 \wedge \dots \wedge x_r) = d(f)u$. In other words, the general alternating map $a: M^r \to N$ may be represented as the alternating map d, times some element of N.

If f, g are two endomorphisms of M, then from the functoriality of Λ^r we see that $d(fg) = d(f)d(g)$.

Note that for $f: M \to M$, the induced map $\Lambda^n f$ multiplies *every* element of $\Lambda^n M$ by the scalar $d(f)$, and that this property uniquely determines that scalar. Hence, though we needed the existence of a basis of r elements to show the existence of such a scalar, the value of the scalar can be characterized independent of choice of basis. This shows why the determinant of an endomorphism of a free module makes sense independent of basis.

Given a linear map $f: M \to N$ between free k-modules of finite rank, how will the induced map Λf act on the general element of ΛM? Let $x_1, \dots, x_r$ be a basis of M, $y_1, \dots, y_s$ a basis of N, and $((f_{ij}))$ the matrix of f (considered to act on the right) with respect to these bases. A typical member of our basis of ΛM has the form

$$p \;=\; x_{i_1} \wedge \dots \wedge x_{i_m} \in \Lambda^m M \qquad (1 \leq i_1 < \dots < i_m \leq r),$$

and a typical member of our basis of ΛN the form

$$q \;=\; y_{j_1} \wedge \dots \wedge y_{j_n} \in \Lambda^n N \qquad (1 \leq j_1 < \dots < j_n \leq s).$$

I claim that the coefficient of q in $(\Lambda f)(p)$ is zero unless $m = n$, and that in this case, it is the determinant of the $n \times n$ submatrix of $((f_{ij}))$ gotten by intersecting the $i_1, \dots, i_n$th rows of that matrix with the $j_1, \dots, j_n$th columns. The first part of the claim follows from the fact that Λf is put together out of maps $\Lambda^n f$, taking $\Lambda^n M$ to $\Lambda^n N$, but has no components carrying $\Lambda^m M$ to $\Lambda^n N$ for $m \neq n$. The second can be seen by composing f on the one side with the inclusion in M of the free submodule M' spanned by $x_{i_1}, \dots, x_{i_n}$, and on the other side with the projection of N onto the free submodule N' spanned by $y_{j_1}, \dots, y_{j_n}$. It is easy to see that on application of Λ^n, the first of these maps sends the single generator $p' = x_{i_1} \wedge \dots \wedge x_{i_n}$ of $\Lambda^n M'$ to the basis element p of $\Lambda^n M$ (also formally $x_{i_1} \wedge \dots \wedge x_{i_n}$, but here distinguished symbolically because it represents an element of a different algebra), and that the second sends the basis element q of $\Lambda^n M$ to the generator $q' = y_{j_1} \wedge \dots \wedge y_{j_n}$ of $\Lambda^n N'$, and all other basis elements to 0. Now the composite f' of f with these maps on the indicated sides is represented by the $n \times n$ submatrix of $((f_{ij}))$ in question, hence $\Lambda^n f'$ sends p' to q' times the determinant of that matrix, hence that determinant is the coefficient of q in $(\Lambda f)(p)$, as desired

If M is a free module with basis $x_1, \dots, x_r$, note that $\Lambda^{r-1} M$ is also free of rank r: for each $i \leq r$ we get a basis vector of this module by *leaving out* the ith factor of $x_1 \wedge \dots \wedge x_r$. Let f be an endomorphism of M. Then the coefficients in the action of $\Lambda^{r-1} f$ with respect to this basis are

essentially the entries of the *classical adjoint* of the matrix of f (Hungerford [**2** p. 353]; written $\widetilde{A}$ but not named in Lang [**3** p.518]). Precisely, the classical adjoint is the *transpose* of a matrix having these entries, with certain changes of sign. Let us now derive in our present context the result that the product of a square matrix with its classical adjoint is the identity times the determinant of the matrix. We will find it easiest to keep track of what should map into what if we start with a linear map $f\colon M \to N$, where M and N are possibly distinct free k-modules of the same rank r, and only at the end specialize to the case where they are the same.

Let us note that every element $x \in M$ induces a map $x\wedge -\colon \Lambda^{r-1}M \to \Lambda^r M$. If we let x run over our basis of M, then, up to sign, these maps run over the obvious basis of $\mathrm{Hom}(\Lambda^{r-1}M, \Lambda^r M)$. To get the correct signs, let us take a slightly modified basis of $\Lambda^{r-1}M$ consisting of the elements

$$\begin{gathered} x_2 \wedge \ldots \wedge x_r, \\ -x_1 \wedge x_3 \wedge \ldots \wedge x_r, \\ \ldots \\ \pm x_1 \wedge \ldots \wedge x_{r-1}, \end{gathered}$$

using a $+$ or $-$ according to whether, on left multiplying by the appropriate x_i, this has to pass through an even or an odd number of factors to reach its proper place, i.e., according as i is odd or even. We can now say that with respect to our given basis of M, the above basis of $\Lambda^{r-1}M$, and the singleton basis $\{x_1, \ldots, x_r\}$ of $\Lambda^r M$, the ''multiplication'' map $M \to \mathrm{Hom}(\Lambda^{r-1}M, \Lambda^r M)$ is represented by the identity matrix.

Now consider the diagram

$$\begin{array}{ccc} M & \xrightarrow{\qquad f \qquad} & N \\ \downarrow & & \downarrow \\ \mathrm{Hom}(\Lambda^{r-1}M,\ \Lambda^r M) & & \mathrm{Hom}(\Lambda^{r-1}N,\ \Lambda^r N) \\ (\Lambda^r f)\circ - \big\downarrow & & \big\downarrow - \circ (\Lambda^{r-1} f) \\ \mathrm{Hom}(\Lambda^{r-1}M,\ \Lambda^r N), & & \end{array}$$

where the upper two vertical arrows are those just described, while the lower ones represent post- and pre-composition with $\Lambda^r f$ and with $\Lambda^{r-1}f$ respectively. The diagram commutes: If we start with an element $x \in M$ we get

$$\begin{array}{ccc} x & \longmapsto & f(x) \\ \downarrow & & \downarrow \\ x\wedge - & & f(x)\wedge - \\ \downarrow & & \downarrow \\ (\Lambda^r f)(x\wedge -) & = & f(x)\wedge (\Lambda^{r-1} f)(-), \end{array}$$

with equality at the bottom because Λf is a ring homomorphism.

Now suppose $M = N$, and we fix a basis of this module, so that f, the top arrow of this diagram, an endomorphism of M, is represented by a certain matrix with respect to our basis. The upper vertical arrows are the same map, which, as we have noted, is represented in the appropriate basis by the identity matrix; the left hand lower vertical arrow is just the identity matrix times $\det(f)$, while the lower right-

hand matrix has for entries the determinants of $(r-1)\times(r-1)$ minors of f, but with signs modified because of our choice of basis of $\Lambda^{r-1}M$, and transposed, because we are letting it act on the first (contravariant) argument of $\mathrm{Hom}(\Lambda^{r-1}M, \Lambda^r N)$. It is easy to verify that it is precisely the classical adjoint of f, so the commutativity of the above diagram shows that the matrix of f followed by its classical adjoint gives $(\det f)\,I$.

I do not see as direct a way to get the corresponding result for composition of these matrices in the opposite order. However, it can be obtained by applying the above result to the dual of f, and taking the transpose of both sides.

Subsequent sections of this note will, like earlier sections, look at properties of exterior algebras belonging to frameworks that also apply to tensor algebras and symmetric algebras. For further results on exterior algebras in particular (and the related construction of Clifford algebras), see Lang [**3**, Chapter XIX].

3. Geometric interpretations of exterior powers. [To be written up some time in the future. Cross-products; differential forms; closed polyhedra]

4. Graded algebras.

Definition 6. *A* graded (*or more precisely,* **Z**-graded) *k-algebra will mean a k-algebra R given with a direct sum decomposition into k-submodules*

$$R = \oplus_{n\in \mathbf{Z}}\, R_n$$

such that

$$1 \in R_0$$

and such that for all m, n,

$$R_m R_n \subseteq R_{m+n}.$$

If R is a graded algebra, an element is said to be homogeneous of degree *n if it belongs to the summand R_n,* homogeneous *if it belongs to one of these summands, which are called the* homogeneous components *of R.*

If R and S are graded algebras, then a homomorphism $R \to S$ as graded algebras means a homomorphism as algebras, which for each n carries R_n into S_n.

(*More generally, for any monoid G, which we will assume written multiplicatively, a G-graded k-algebra means a k-algebra R given with a direct sum decomposition into submodules*

$$R = \oplus_{g\in G}\, R_g$$

such that

$$1 \in R_e$$

and for all $g, h \in G$

$$R_g R_h \subseteq R_{gh}.$$

Homogeneous components and graded-algebra homomorphisms are defined in the same way for this case.)

We thus see that $k\langle M\rangle$, ΛM and SM can all be regarded as **Z**-graded k-algebras, or, indeed, as

N-graded k-algebras, where **N** is the additive monoid of nonnegative integers.

Gradings constitute an important tool in both commutative and noncommutative ring theory. We will not be able to look at that subject in this course; we have introduced the definition only so that we can define the concept of the next section, which is relevant to exterior algebras.

5. ''Graded'' or ''super'' or ''topologists''' commutativity – a sketch. It is easy to verify that in an exterior algebra, homogeneous elements of even degree commute with everything, while homogeneous elements of odd degree anticommute with each other. In other words, for $a \in \Lambda^m M$, $b \in \Lambda^n M$,

$$ab = (-1)^{mn} ba.$$

This same equation turns out to hold frequently in the cohomology rings studied in topology, so that some topologists have taken to calling this ''commutativity'' as graded rings!

In recent years, the study of graded rings with this property has flourished under the name ''super'' commutativity. In this context, analogs of a large number of familiar concepts can be defined. For instance, just as in a not necessarily commutative algebra, one can define commutator brackets, which measure the failure of commutativity, and then define objects with an operation which behaves like commutator brackets, called ''Lie algebras'', so in a graded ring, one can define ''super'' commutator brackets, which act on homogeneous elements by the law $[a, b] = ab - (-1)^{\deg(a)\deg(b)} ba$, and define ''super Lie algebras'' as graded k-modules with an operation satisfying the identities of these brackets. And just as one may define the tensor product of two algebras, with multiplication based on the formula $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$, which, if the algebras are commutative, is itself commutative and gives their coproduct in the category of commutative algebras, so one may define a modified multiplication on the module-theoretic tensor product of two graded algebras using the formula $(a \otimes b)(a' \otimes b') = (-1)^{\deg(b)\deg(a')} aa' \otimes bb'$ for homogeneous elements a, b, a', b', and this turns out to preserve super commutativity, and to give the coproduct in the class of such algebras.

The condition of super commutativity is slightly weaker than the alternating condition imposed in the exterior algebra construction. If one takes $b = a$ in the equation $ba + ab = 0$, one gets $2a^2 = 0$, which, if 2 is invertible in the base ring k, is indeed equivalent to the alternating condition; but is not in cases such as $k = \mathbf{Z}$ or $\mathbf{Z}/2\mathbf{Z}$. So over a ring in which 2 is invertible, the exterior algebra construction, regarded as a functor from k-modules M to super commutative k-algebras, is left adjoint to the functor taking such an algebra to its homogeneous component of degree 1; and the functor associating to a set X the exterior algebra on the free k-module on X is left adjoint to the functor taking a super commutative k-algebra to the *set* of its elements of degree 1; i.e., this exterior algebra is a super commutative k-algebra ''freely generated by an X-tuple of elements of degree 1''. But over a general base-ring k, though such left adjoints still exist, they are somewhat larger than the exterior algebras.

Note that in defining ''super commutativity'', one only needs to know whether degrees are odd or even, not their exact values. Thus, the concept can be defined on $\mathbf{Z}/2\mathbf{Z}$-graded rings. Moreover, any **Z**-graded ring can be regarded as a $\mathbf{Z}/2\mathbf{Z}$-graded ring by calling the sum of the components R_n with n even the ''even'' component and the sum of those with n odd the ''odd'' component. Thus, the concept of super commutativity is often defined for $\mathbf{Z}/2\mathbf{Z}$-graded rings, with the understanding that **Z**-graded rings are automatically regarded as $\mathbf{Z}/2\mathbf{Z}$-graded in this way.

Now suppose we define a ''forgetful functor'' taking each super commutative $\mathbf{Z}/2\mathbf{Z}$-graded k-algebra R to the ordered pair of sets (R_0, R_1). This functor has a left adjoint, which takes a pair of sets (X_0, X_1) to the super commutative k-algebra ''freely generated by an X_0-tuple of even elements and an X_1-tuple of odd elements''. Assuming 2 is invertible in k, we see that when $X_0 = \emptyset$, this is the

exterior algebra on X_1; when $X_1 = \emptyset$, it is the polynomial algebra on X_0, while for general X_0 and X_1, it is the tensor product of these two algebras.

6. Direct sum decompositions. In Theorem 5 we described the form our constructions take when applied to free modules. Let us consider what we get when we start with a general direct sum of modules, $M \oplus N$.

The tensor power construction is based on iterating the pairwise tensor product construction, which satisfies the laws $(A \oplus B) \otimes C \cong (A \otimes C) \oplus (B \otimes C)$ and $A \otimes (B \oplus C) \cong (A \otimes B) \oplus (A \otimes C)$. It is not hard to deduce that $\otimes^n(M \oplus N)$ is the direct sum of all n-fold tensor products of strings of M's and N's (2^n summands in all).

The cases of the symmetric and exterior powers are not so obvious. Note that $\Lambda^n(M \oplus N)$, respectively $S^n(M \oplus N)$, is spanned as a k-module by products of elements of M and elements of N, and that by the anticommutativity or commutativity law, every such product may be reduced to a product in which all factors coming from M precede all factors coming from N. Within each of these two strings of factors, the commutative or alternating law can be used to get further equalities, but it does not appear that there will be any relations connecting the two strings of elements, except those arising from bilinearity. This motivates the next result, for which I will briefly sketch three methods of proof.

Theorem 7. *Let M and N be k-modules. Then*

$$\Lambda^p(M \oplus N) \cong \oplus_{m+n=p} (\Lambda^m M) \otimes (\Lambda^n N),$$

via the map

$$x_1 \wedge \ldots \wedge x_m \wedge y_1 \wedge \ldots \wedge y_n \longleftarrow (x_1 \wedge \ldots \wedge x_m) \otimes (y_1 \wedge \ldots \wedge y_n),$$

and likewise

$$S^p(M \oplus N) \cong \oplus_{m+n=p} (S^m M) \otimes (S^n N),$$

via the map

$$x_1 \ldots x_m y_1 \ldots y_n \longleftarrow (x_1 \ldots x_m) \otimes (y_1 \ldots y_n).$$

Sketch of 3 proofs. I will discuss the exterior-algebra case. The symmetric-algebra case is similar.

Method 1. The idea is to construct a k-multilinear alternating map from $(M \oplus N)^p$ to the indicated direct sum, and verify that it has the universal property of p-fold exterior multiplication. The map is defined so that a string consisting of m elements from M and n elements from N in some order goes to the element of $(\Lambda^m M) \otimes (\Lambda^n N)$ gotten by rearranging them so that all members of M precede all members of N, multiplying them together in the respective exterior powers, and putting on a + or – sign according to whether the ''rearrangement'' involved an odd or an even number of transpositions. This is then extended by multilinearity to terms consisting of sums of elements of M and elements of N. Once this operation is defined precisely, it is straightforward to establish that it has the desired universal property.

Method 2. Use the version of the diamond lemma in [**1**, §6] which deals with presentations of rings in terms of generating (bi)modules instead of elements. Present $\Lambda(M \oplus N)$ in terms of the modules M and N, get a normal form for this algebra, and extract $\Lambda^p(M \oplus N)$ as the pth power therein of the generating module $M \oplus N$.

Method 3. It is easy to verify by universal properties that $\Lambda(M \oplus N)$ is the coproduct, in the category of ''super commutative'' graded k-algebras, of ΛM and ΛN. (Key point: an easy computation shows

that in a super commutative algebra, a sum of odd-degree elements with zero square again has zero square.) Now use the description of coproducts in this category as tensor products, referred to in the preceding section. The degree-p component of $(\Lambda M) \otimes (\Lambda N)$ is $\oplus_{m+n=p} (\Lambda^m M) \otimes (\Lambda^n N)$, as desired. □

7. Functorial notes. Given a k-module M and a nonnegative integer n, the universal property of each of $\otimes^n M$, $\Lambda^n M$, $S^n M$ can be considered as saying that the module is a *representing object* for a certain set-valued functor. E.g., $\Lambda^n M$ represents the functor which takes every module P to the set of alternating k-multilinear maps $M^n \to P$. Now it frequently happens that when, for each object X of a category $\mathbf{C}$, we can construct an object $F(X)$ of a category $\mathbf{D}$, representing a covariant functor $\mathbf{D} \to \mathbf{Set}$ defined in terms of X, then these objects fit together to form a functor $F\colon \mathbf{C} \to \mathbf{D}$, which is a left adjoint to some functor $U\colon \mathbf{D} \to \mathbf{C}$. In the present case each of our three module constructions does indeed give a functor k-**Mod** $\to$ k-**Mod**, but we have an exception to the second general observation: none of these functors is a left adjoint. Indeed, left adjoint functors respect coproducts, but the observations of the preceding section show that for $n > 1$, none of these functors takes direct sums to direct sums.

Curiously, however, when we put these modules together to form k-algebras, one of the resulting constructions, the tensor algebra, does, as we have seen, become a left adjoint – the left adjoint to the forgetful functor from k-algebras to k-modules. The other two are still not adjoints when regarded as functors to general k-algebras, but as we have also seen, the symmetric algebra construction becomes one when regarded as a functor to commutative k-algebras. It is not hard to show that, similarly, the exterior algebra functor, when regarded as a functor to ''super-commutative'' graded k-algebras in which odd-degree elements satisfy $a^2 = 0$, becomes left adjoint to the ''module of elements of degree 1'' functor.

On the other hand, all of our constructions of both modules and algebras can be made into left adjoints if their descriptions are adjusted somewhat. For example, if the n-fold tensor power construction is inflated so that the codomain category is not the category of k-modules, but the category of 3-tuples (M, a, P) where M and P are k-modules and a is a k-multilinear map $M^n \to P$, then it becomes a left adjoint to the forgetful functor $(M, a, P) \mapsto M$.

8. Operations relating these algebras and their duals.

[To be written up some time in the future?]

For more on the exterior algebra cf. Lang's *Algebra*, 3rd edition, Chapter XIX.

REFERENCES

1. George M. Bergman *The diamond lemma for ring theory,* Advances in Mathematics **29** (1978) 178-218. MR **81b**:16001.

2. Thomas W. Hungerford, *Algebra,* Graduate Texts in Mathematics, vol. 73, Springer-Verlag, 1974.

3. Serge Lang, *Algebra,* Addison-Wesley, (1st ed. 1965, 2nd ed. 1984) 3rd ed. 1993.